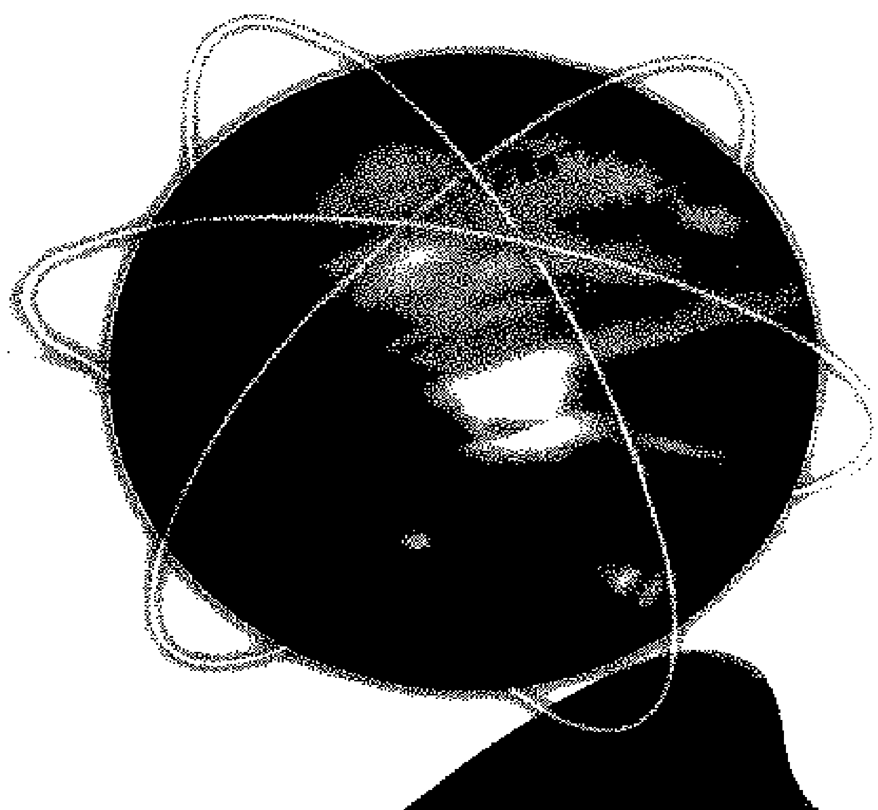
سلسلة الرضا للإنترنت 3

البحث عن المعلومات في الإنترنت

Internet Information Retrieval

د. عمار خير بك



دار الرضا للنشر

البحث عن المعلومات في الإنترنت

الدكتور عمّار خيربك

سلسلة الرضا للإنترنت -٣-

سلسلة الرضا للمعلومات ٦٠

سلسلة علمية متميزة لنشر ثقافة الإدارة الحديثة والمعلوماتية بغية تطوير المؤسسات والشركات التي تسعى للريادة.

دار الرضا للنشـــر

تجهيـز – قـرب فندق برج الفـردوس – هـاتف: ٢٢٢٤٦١٧
تلفاكس: ٢٢٢٢١٦٣
ص.ب: ٤٢٦٧
E-mail: Reda-Center @ net.sy

اشراف على سلسلة الانترنت: د. عمّار خير بك
د. راضي خازم

نيسان ٢٠٠٠

مقدمة الكتاب

بدأت قصتي مع البحث عن المعلومات ونظم الوسائط الفائقة (Hypermedia Systems) منذ حوالي عشر سنوات، وتحديداً في أيار 1990. ففي ذلك الوقت لم يكن قد مضى على تخرجي من هندسة المعلوماتية (في المعهد العالي للعلوم التطبيقية والتكنولوجيا) سوى سبعة أشهر، وكنت أعمل في مشروع بهدف الحصول على دبلوم سمي آنذاك دبلوم دراسات عليا (DES). كان عنوان هذا المشروع /توليد قاعدة بيانات متعددة الوسائط عن المواقع الأثرية في سورية - تطبيق على مدينة تدمر الأثرية/، وكانت بيئة العمل هي نظام HyperCard على حواسيب Macintosh. لم أكن أعلم آنذاك أن هذا النظام هو أول نظام معلوماتي فعلي لتوليد الوسائط الفائقة، وكنت أستمتع جداً بالعمل فيه وخصوصاً بإمكانات التخاطب العالية التي يقدمها، ولغة البرمجة القوية Hypertalk التي أتاحت لي التقدم بسرعة في إنجاز المشروع.

في هذا الوقت جاءنا بروفيسور فرنسي زائر اسمه ايف كيارميلا (Yves Chiaramella) ليعطينا مادة سُميت /البحث عن المعلومات/ خلال أسبوعين فقط. كانت هذه المادة جديدة علينا كل الجدة من حيث المحتوى وطريقة العرض والأسئلة الكبيرة التي تطرحها. وأثناء إقامة البروفيسور كارميلا اطلع على مشروعي وتوقف عنده طويلاً، وسألني أسئلة كثيرة منها ما استطعت الإجابة عليها ومنها ما لم استطع بسبب الطبيعة الفلسفية لأسئلته. قال لي قبل مغادرته سوريا ما ملخصه أني أستطيع العمل في هذا المجال تحت إشرافه في جامعة جوزيف فورييه في غرنوبل/فرنسا، لأن البحث العلمي في موضوع الوسائط الفائقة كان في مرحلة الإقلاع هنالك.

وصلت إلى فرنسا في منتصف تشرين أول من عام 1990 متأخراً ثلاثة أسابيع عن بدء الدراسة في دبلوم الدراسات المعمقة (DEA)، فذهبت مباشرة إلى البروفيسور كارميلا لأجده قد وفى بوعده إذ كان المشروع بانتظاري تحت عنوان /مكاملة نظم استرجاع (البحث عن) المعلومات مع نظم الوسائط الفائقة/ في المعهد الوطني للبوليتكنيك في غرنوبل. وكم كانت سعادتي كبيرة عندما اكتشفت أن ايف هو مدير أهم مختبر للمعلوماتية (مختبر هندسة المعلوماتية LGI) في الـ IMAG (معهد المعلوماتية والرياضيات التطبيقية في غرنوبل)، وأنه يعتبر المرجع الأول في مجال البحث عن المعلومات في فرنسا والثالث على مستوى العالم.

تابعت العمل في نفس الموضوع في الدكتوراه، وكانت أيام البحث العلمي في فرنسا حلماً جميلاً بسبب المناخ العلمي والاجتماعي الرائعين هناك، وبسبب الفرص العديدة التي أتيحت لي بحضور مؤتمرات عالمية وبإلقاء محاضرات في مجال البحث عن المعلومات في أكثر الأجواء العلمية رصانة في أوربة الغربية والولايات المتحدة الأمريكية.

أثناء تواجدي في فرنسا أيضاً عاصرت انفجار واحدة من أهم ثورات القرن العشرين، وهي تلك المتمثلة بشبكة الإنترنت. لقد أحسست منذ اللحظة الأولى أني معنيٌّ بها إلى أقصى درجة وخصوصاً بعد ظهور ما سمي بالشبكة العنكبوتية العالمية (WWW أو الوب اختصاراً) باعتبار أنها ترتكز ارتكازاً كاملاً على مفاهيم الوسائط الفائقة وبعض مفاهيم البحث عن المعلومات. لذلك فقد صممت صفحة موطن (Home Page) عن سورية في صيف 1993، أي بعد ظهور أول متصفح (Browser) على الشبكة وكان اسمه موزاييك بشهرين. من أبرز ما قمت به في تلك الفترة هو المشاركة بالمؤتمر العالمي الأول عن الـ WWW (في أيار 1994 – جنيف – سويسرا)، ولعلّ أهم ما أتذكره من ذلك المؤتمر يتلخص بأمرين اثنين: التعرّف على الباحث الإنكليزي تيم برنرز لي (Tim Berners-Lee) مخترع الـ WWW، والمؤامرة البيضاء التي رتبتها مع صديقي الفرنسي سيرج رفيرول (Serge Rouvyrol) أثناء المؤتمر، حيث كانت لنا مداخلة عن ضياع المستخدم في فضاء معلومات الوب، وملخصها أن يبدأ المستخدم ببحث عن معلومات معينة وينتهي عند معلومات أخرى مختلفة جذرياً، وقد كانت النتيجة عرض صفحة موطن سورية على شاشة المؤتمر.

بعد أن عدت إلى سورية دأبت، كلما سنحت لي الفرصة، علـى إلقاء محـاضرة أو المشـاركة في ورشة عمل أو الإشراف على مشاريع طلابية في موضوع البحث عن المعلومات عموماً، و في موضـوع البحـث عن المعلومات في الإنترنت خصوصاً. وتبينت لي، بسبب الاهتمام بـهذه المحـاضرات وبسبب انتشار استخدام الإنترنت، ضرورة كتابة كتاب عن هذا الموضوع ليصبح في متناول الجميع، وكـانت النتيجـة هذا الكتاب الذي بين أيديكم.

لا أدعي أني أعددت كتاباً متكاملاً عن الموضوع، وإنما أعتبر هذا الكتاب خطوةً أولى في هـذا الطريـق. لا شكّ أن هذا الكتـاب سيتطور في المستقبل القريب في طبعـات متلاحقـة تعتمد على ردود أفعـال وملاحظات القراء عليه، وعلى التطور العلمي المتسارع في هذا المجال.

وفي الختام أردّد هذه القولة المشهورة للعماد الأصفهاني (نقلاً عن معجم المورد):

"إني رأيت أنه لا يكتب أحدٌ كتاباً في يومه إلاّ قال في غده:

لو غيّر هذا لكان أحسن، ولو زيد هنا لكان يستحسن،

ولو قُدّم هذا لكان أفضل، ولوترك هذا لكان أجمل.

وهذا من أعظم العبر،

وهو دليل على استيلاء النقص على جملة البشر."

د. عمّار خيربك

دمشق في 2000/3/1

فهرس المحتويات

مقدمة الكتاب ________ 1

فهرس المحتويات ________ 5

الجزء الأول: خلفية نظرية عن نظم البحث عن المعلومات ________ 11

مقدمة الجزء الأول ________ 13

الفصل الأول: تأطير إشكالية البحث عن المعلومات ________ 15

1-1 مقدمة ________ 15

1 -2 تأثير التطور التكنولوجي على مسألة البحث عن المعلومات ________ 16

1-3 العوائق الجديدة أمام البحث عن المعلومات ________ 19

1-4 التصنيف العريض لنظم البحث عن المعلومات ________ 20

1-5 خلاصة ________ 22

الفصل الثاني: نظم البحث عن المعلومات ________ 23

2-1 مقدمة ________ 23

2-2 البحث عن المعلومات في نظم قواعد البيانات ________ 24

2-3 البحث عن المعلومات في نظم استرجاع المعلومات ________ 25

2-3-1 تعريف نظم استرجاع المعلومات 25
2-3-2 المخطط العام لنظم استرجاع المعلومات 26
2-3-3 تقنيات نظم استرجاع المعلومات 29
2-3-4 عرض لبعض نظم استرجاع المعلومات 36
2-4 البحث عن المعلومات في نظم الوسائط الفائقة 39
2-4-1 مقدمة 39
2-4-2 تعريف الوسائط الفائقة 42
2-4-3 بنية نظم الوسائط الفائقة 43
2-4-4 بعض إشكاليات نظم الوسائط الفائقة 46
2-4-5 عرض لبعض نظم الوسائط الفائقة 48
2-4-6 اعتبارات بناء نظام وثائق فائقة والتقنيات المتبعة 57
2-4-7 البحث عن المعلومات 58
2-5 البحث عن المعلومات في النظم الهجينة 59
الفصل الثالث: تقييم ومستقبل البحث عن المعلومات 61
3-1 تقييم نظم استرجاع المعلومات 61
3-2 تقييم نظم الوسائط الفائقة 63
3-3 بعض التصورات المستقبلية لنظم استرجاع المعلومات 64
3-4 بعض التصورات المستقبلية لنظم الوسائط الفائقة 64
3-5 مقارنة نظم استرجاع المعلومات مع نظم الوسائط الفائقة 65
3-6 مستقبل نظم البحث عن المعلومات 66

الجزء الثاني: طرق البحث عن المعلومات في الإنترنت 67

مقدمة الجزء الثاني 69

الفصل الرابع: شبكة الإنترنت وخدمات البحث عن المعلومات 71

4-1 تعريف شبكة الإنترنت 71

4-2 تاريخ الإنترنت 72

4-3 خدمات البحث عن المعلومات في الإنترنت 76

4-3-1 الشبكة العنكبوتية العالمية WWW 76

4-3-2 غوفر Gopher 80

4-3-3 فيرونيكا Veronica 82

4-3-4 البحث في مواقع FTP بواسطة Archie 84

4-3-5 البحث في النصوص بواسطة WAIS 86

الفصل الخامس: محركات البحث وطرق تقييمها 89

5-1 تعريف محركات البحث 89

5-2 معايير تقييم محركات البحث 91

5-2-1 البحث البسيط 91

5-2-2 البحث المتقدم 94

5-2-3- الفلترة 95

5-2-4 تصنيف المواضيع 96

5-2-5 التغذية الراجعة 96

5-2-6 التحكم بالإظهار 96

5-2-7 السرعة 97

5-2-8 حجم قاعدة البيانات 97

5-2-9 عوامل أخرى ____ 99

الفصل السادس: محركات البحث العامة ____ 99

6-1 محرك البحث Yahoo! ____ 101
6-1-1 واجهة Yahoo! ____ 103
6-1-2 البحث عن المعلومات في Yahoo! ____ 117
6-1-3 خدمات المعلومات الأخرى ____ 120
6-1-4 إضافة موقع إلى محرك Yahoo! ____ 120
6-1-5 الانضمام إلى مجتمع Yahoo! ____ 120
6-2 محرك البحث HotBot ____ 122
6-2-1 واجهة HotBot ____ 123
6-2-2 البحث عن المعلومات في HotBot ____ 124
6-2-3 خدمات المعلومات الأخرى ____ 136
6-2-4 إضافة موقع إلى محرك البحث HotBot ____ 137
6-3 محرك البحث AltaVista ____ 139
6-3-1 واجهة AltaVista ____ 140
6-3-2 البحث عن المعلومات في AltaVista ____ 141
6-3-3 خدمات المعلومات الأخرى ____ 156
6-3-4 إضافة موقع إلى محرك البحث AltaVista ____ 157
6-4 محرك البحث InfoSeek ____ 159
6-4-1 واجهة InfoSeek ____ 159
6-4-2 البحث عن المعلومات في InfoSeek ____ 161
6-4-3 خدمات المعلومات الأخرى ____ 171
6-4-4 إضافة موقع إلى محرك البحث InfoSeek ____ 172
6-5 محرك البحث Lycos ____ 173

6-5-1 واجهة Lycos ____ 173
6-5-2 البحث عن المعلومات في Lycos ____ 175
6-5-3 أدوات المعلومات الأخرى ____ 182
6-5-4 إضافة موقع إلى محرك البحث Lycos ____ 183
6-6 محرك البحث Excite ____ 183
6-6-1 واجهة Excite ____ 184
6-6-2 البحث عن المعلومات في Excite ____ 185
6-6-3 إضافة موقع إلى محرك البحث Excite ____ 194
6-7 محرك البحث WebCrawler ____ 194
6-7-1 واجهة WebCrawler ____ 195
6-7-2 البحث عن المعلومات في WebCrawler ____ 196
6-7-3 أدوات المعلومات الأخرى ____ 200
6-7-4 إضافة موقع إلى محرك البحث WebCrawler ____ 201

الفصل السابع: محركات البحث البينية والعربية ____ 203

7-1 محركات البحث البينية ____ 203
7-2 محرك البحث البيني MetaCrawler ____ 204
7-2-1 واجهة MetaCrawler ____ 205
7-2-2- البحث عن المعلومات MetaCrawler ____ 207
7-2-3 إضافة موقع إلى محرك البحث MetaCrawler ____ 213
7-3 محرك البحث البيني The Big Hub ____ 214
7-3-1 واجهة theBigHub ____ 214
7-3-2 البحث عن المعلومات في theBigHub ____ 216
7-3-3 أدوات المعلومات الأخرى ____ 217
7-4 محركات البحث العربية ____ 217
7-5 محرك البحث العربي أين/نظرة ____ 217

الفصل الثامن: مقارنة بين محركات البحث ______ 219
8-1 البحث حسب المحتوى ______ 220
8-2 استخدام التصنيف بالمجلدات ______ 224
8-3 تسجيل المواقع ضمن محركات البحث ______ 226
8-4 شعبية محركات البحث ______ 228
8-5 أهمية محركات البحث ______ 230
8-6 أزمنة استخدام محركات البحث ______ 232
الفصل التاسع: خلاصة وتوصيات ______ 233
9-1 بعض النصائح للباحثين عن المعلومات في الإنترنت ______ 233
9-2 بعض التصورات المستقبلية للبحث عن المعلومات في الإنترنت ______ 234
ملحق: بعض العناوين المفيدة في البحث عن المعلومات في الإنترنت ______ 235
1. محركات البحث العامة ______ 235
2. محركات البحث المخصصة للأطفال ______ 238
3. محركات البحث البينية ______ 239
4. محركات البحث عن الوسائط المتعددة ______ 241
5. محركات البحث الإقليمية ______ 242
6. محركات البحث عن الأخبار ______ 243
المراجع المستخدمة في الكتاب ______ 245

الجزء الأول

خلفية نظرية لنظم البحث عن المعلومات

مقدمة الجزء الأول

قد يبدو للوهلة الأولى أن هذا الجزء مقحمٌ في الكتاب، أو أن له طابعاً أكاديمياً لا يتناسب مع هدف الكتاب، في حين سيكتشف القارئ بسرعة أن هذا الجزء سيساعده مساعدةً كبيرة في فهم الجزء الثاني.

فالهدف من وجود هذا الجزء، هو تعريف القارئ على نحوٍ مبسط وسريع بمجالات علمية شبه مجهولة في أوساطنا العلمية كنظم الوسائط الفائقة ونظم استرجاع المعلومات، رغم الحاجة الماسة لها في كثير من القطاعات كالمكتبات العامة ومراكز التوثيق والموسوعات والمراكز الأكاديمية كالجامعات ومراكز البحوث. يتيح هذا الجزء أيضاً المجال أمام القارئ للتعرف على أهم الخلفيات النظرية التي بُنيت عليها أدوات البحث عن المعلومات في الإنترنت (موضوع هذا الكتاب). ففهم الخلفية النظرية لمحركات البحث (Search Engines) مثلاً، يساعد على التعامل مع محركات البحث تعاملاً فعّالاً، كما يسمح بتفسير بعض النتائج الغريبة التي يمكن أن تحدث عند استخدام هذه المحركات.

سنسعى في الفصل الأول لتأطير مشكلة البحث عن المعلومات من خلال شرح أبعادها وحدودها، والعوامل المؤثرة بها، إضافةً إلى عرض بعض معوقات البحث، ثم التصنيف العريض لنظم البحث عن المعلومات.

ننتقل في الفصل الثاني إلى شرح مختلف أنواع نظم البحث عن المعلومات شرحاً نظرياً مبسطاً مركزين الاهتمام في النظم غير المعروفة معرفةً جيدة بالنسبة للقارئ الكريم. نتعرض إذن لنظم البحث عن المعلومات في قواعد البيانات، ونظم استرجاع المعلومات، ونظم الوسائط الفائقة، والنظم الهجينة التي تتألف من اجتماع معظم وظائف نظم البحث السابقة.

لا شكّ بأن أي أداة للبحث عن المعلومات في الإنترنت (كما سنرى في الجزء الثاني من هذا الكتاب) تعتمد اعتماداً جزئياً أو كلياً على واحد أو أكثر من النظم المشروحة في هذا الفصل.

سنتعرض أخيراً، في الفصل الثالث، بإيجاز لتقييم هذه النظم ومقارنتها بعضها ببعض، ولمستقبل هذه النظم وآفاق تطورها من وجهة نظرنا الشخصية.

وربما لا بد من التنويه أن جزئي الكتاب منفصلان إلى حدّ كبير، ويمكن للقارئ المستعجل لاستكشاف البحث عن المعلومات في الإنترنت أن يتجاوز الجزء الأول. في حين يحتاج القارئ المهتم بفهم الأبعاد الحقيقية لآليات البحث عن المعلومات إلى دراسة هذا الجزء دراسةً عميقة.

الفصل الأول

تأطير إشكالية البحث عن المعلومات

1-1 مقدمة

تهدف مختلف أنظمة البحث عن المعلومات عموماً إلى "البحث في قواعد معلومات ضخمة عن وثائق متعددة الوسائط (Multimedia Documents) (نصوص، صور، أصوات، فيديو، ...) ملائمة لحاجة معينة لدى المستخدم بطرق فعالة تتطلب أقل ما يمكن من الجهد والوقت".

تسمح بعض الأنظمة بإجراء عمليات البحث حسب صفات للمعلومات (Attributes) مخزنة سلفاً وفق بنية محددة. كأن نبحث مثلاً عن لائحة مرتبة أبجدياً بأسماء الأشخاص في مدينة دمشق الذين تتجاوز أعمارهم 50 سنة والذين يعملون في مهنة النقش على النحاس، أو أن نطبع لائحة بأسماء المستودعات التي تحتوي أقل من 100 مادة مختلفة.

تتيح بعض الأنظمة الأخرى الإجابة على أسئلة من نمط آخر تركّز على مضمون المعلومات (وربما هيكلتها ونوع الوسائط المؤلفة لها) دون الأخذ بعين الاعتبار صفات المعلومات المخزنة في النظام. كأن نسأل مثلاً: "ما هي مجموعة الوثائق التي تتكلم عن الأورام الخبيثة التي تصيب الجهاز الهضمي؟"، أو أبعد من ذلك "ما هي مجموعة الوثائق التي تتكلم عن الأورام الخبيثة التي تصيب الجهاز التنفسي

وتحتوي على صور شعاعية موضّحة ؟"، أو "ما هي مجموعة الصور التي تجسّد انتصارات نابليون إبان فترة حكمه ؟".

بينما تعتمد أنظمة أخرى عن أسلوب سؤال المستخدم عن حاجته من المعلومات وتسمح له بالولوج المباشر إلى فضاء المعلومات، ليقوم بنفسه بالتنقل بين هذه المعلومات وفق الإمكانات المتاحة لهذه الأنظمة حتى الوصول إلى حاجته.

1-2 تأثير التطور التكنولوجي على مسألة البحث عن المعلومات

من الطبيعي أن تتطلب الإجابة على الأسئلة السابقة، أو إتاحة إمكانيات الوصول المباشر والتنقل بين المعلومات، الكثير من المعالجة والذكاء من قبل الآلة. وذلك لم يكن ممكناً في السابق إلى أن ظهرت تطورات بارزة جداً في تكنولوجيا المعلومات والحواسيب والاتصالات أدت إلى انفجار مسألة البحث عن المعلومات. نذكر من هذه العوامل:

أ- تكنولوجيا وسائط التخزين والمعالجات والذواكر

كان لهذا العامل الدور الكبير في بناء أنظمة البحث عن المعلومات، فتخزين الوثائق بحاجة إلى وسائط تخزين ذات سعة كبيرة، كما أن تطوير تقنيات ضغط وتخزين الصوت والصورة سمح بتخزين هذه الوسائط تخزيناً اقتصادياً. من ناحية أخرى، تتصف عمليات البحث والفرز (Searching and Sorting) بأنها ذات تعقيد زمني كبير، لذا نجد أن الأداء يتحسن عندما نستخدم حاسوباً قوياً (معالج + ذاكرة حية) قادراً على القيام بالحسابات بزمن أقصر.

ب- تكنولوجيا التفاعل بين الإنسان والآلة (Human-Machine Interaction)

أدى تطور وسائل الاتصال ودخول الحاسوب في شتى ميادين الحياة على نحو قوي وفعّال، إلى التوجّه نحو استخدام الواجهات التخاطبية الأليفة الاستخدام (Friendly Interfaces) من أجل

تسهيل تعامل المستخدمين مع البرمجيات والنظم الحاسوبية عموماً ونظم البحث عن المعلومات خصوصاً. لذلك فقد طُورت تقنيات جديدة في مجال التفاعل بين الإنسان والآلة بهدف السماح للمستخدمين بأن يكونوا أكثر فعالية أثناء بحثهم عن المعلومات في قواعد المعلومات المؤتمتة.

لقد كانت الواجهات التخاطبية لأنظمة البحث عن المعلومات "القديمة" تجعل استخدام هذه النظم أمراً عقيماً. إذ كان المستخدم يضيع بالأوامر التي يجب أن يتعلمها من أجل الاستخدام ويعاني من صعوبة فهم آلية طرق استجابة هذه النظم على الطلبات التي يكتبها (إن نجح). كما كان المستخدم يغرق في حلقات استفسارية طويلة جداً يُسأل فيها باستمرار عن تحديد حاجته أكثر.

أضف إلى ذلك عدم إمكان تعامل هذه النظم إلاّ مع النصوص وعدم إتاحة إمكانات الانتقال المباشر إلى المعلومات.

ج- تكنولوجيا الوسائط المتعددة (Multimedia) والوسائط الفائقة (Hypermedia)

كان لا بد من التفكير باستراتيجيات جديدة لتقديم المعلومات إلى المستخدم، خاصة بعد انفجار ظاهرة أتمتة المعلومات وتزايد أهمية البحث عنها، وتعدد وسائطها من نصوص إلى صور ومقاطع فيديو ومقاطع صوتية وغيرها. كما دفع التطور في تقنيات التفاعل بين الإنسان والآلة إلى تشجيع هذا الاتجاه. فكانت النتيجة إمكانات الحصول على وثائق متعددة الوسائط تحتوي أنماطاً مختلفة من المعلومات، ثم قاد ذلك إلى تطوير هذه الوثائق من ناحية طريقة تقديمها إلى المستخدم، وإتاحة المجال له للانتقال المباشر إلى المعلومات وتصفحها وصولاً إلى حاجته من المعلومات. لذلك كان الاتجاه نحو استخدام الوسائط الفائقة التي سمحت بطرق عرض جديدة ومميزة أمراً طبيعياً ومطلوباً.

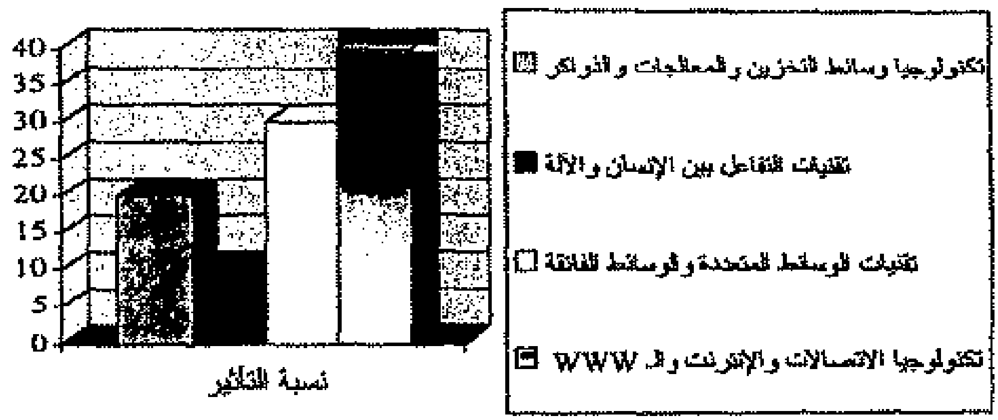
ومن الطبيعي أن يؤدي الاستخدام المشترك لتكنولوجيا الوسائط المتعددة والفائقة مع تكنولوجيا التفاعل بين الإنسان والآلة إلى ظهور جيل جديد من البرمجيات والنظم (بما فيها نظم البحث عن المعلومات) يتصف بالمرونة، وسهولة ومتعة الاستخدام، وتكامل مختلف أنماط المعلومات.

د- تكنولوجيا الاتصالات والإنترنت

لقد أدى التطور في تكنولوجيا الاتصالات إلى انتشار المعلومات انتشاراً كبيراً، وبالتالي إلى الحاجة إلى أنظمة تقوم بالبحث الآلي عن المعلومات. وقد زاد في ذلك ظهور الإنترنت ثم الشبكة العنكبوتية العالمية (World Wide Web <WWW>) عليها، والتي تشكل ذروة الانفجار المعلوماتي في عصرنا هذا. فالعالم أجمع بتنوع معلوماته قد أصبح متاحاً لمختلف المستخدمين في جميع أنحاء العالم من خلال شاشة الحاسوب الصغيرة. وأمام هذا الكم الهائل من المعلومات، يجب التفكير بحل جدي يسمح بالوصول إلى المعلومات التي يطلبها المستخدمون بسرعة وفاعلية.

لقد أزاحت الشبكة العنكبوتية العالمية عبء نشر المعلومات عن كاهل مؤسسات الاتصال لتلقي بالحمل على عاتق منتجي المعلومات الذين وجدوا طريقة مميزة وسهلة الاستخدام لنشر جميع ثقافات وعلوم البشر إلى كل مستخدمي الشبكة بآن واحد ودون تمييز.

يعرض الشكل (1-4) رأينا الشخصي في نسبية تأثير العوامل السابقة في مسألة البحث عن المعلومات.



الشكل (1-4): النسب المئوية لتأثير العوامل الجديدة في انفجار مسألة البحث عن المعلومات

1-3 العوائق الجديدة أمام البحث عن المعلومات

إن إمكانات المعالجة الجديدة والوصول إلى حلول تقنية لا تعني مطلقاً أن مسألة البحث عن المعلومات أصبحت محلولة الآن، لا بل بالعكس فإن التطورات السابقة أدت إلى إفراز عقبات جديدة تضاف إلى العقبات السابقة لتشكل تحديات جديدة لمسألة البحث عن المعلومات.

نذكر من هذه العقبات/ التحديات ما يلي:

أ- وجود كم هائل من المعلومات المتدفقة كل يوم إلى شبكة الإنترنت والتي تحتاج إلى تخزين ومعالجة وإدارة، ويُضاف لها بنوك المعلومات الضخمة السابقة التي أصبحت أيضاً متاحة على الشبكة.

ب- الطلب الزائد على المعلومات بسبب التزايد الكبير والمستمر في أعداد مستخدمي الشبكة (أكثر من 200 مليون مستخدم حتى الآن) وهذا ما يفرض على أنظمة البحث أن تتيح إمكانات الاستخدام لجميع هؤلاء المستخدمين على اختلاف ثقافاتهم ودرجة تعلمهم وحاجاتهم من المعلومات.

ج- طبيعة المعلومات التي أصبحت شديدة التنوع سواء بسبب طرق عرضها وبنيتها (وسائط متعددة وفائقة ومهيكلة وموصّفة ...)، أو بسبب انتمائها إلى مختلف المجالات الممكنة: علمية و اجتماعية و سياسية و اقتصادية و تجارية تسويقية وصولاً إلى مجالات الإرهاب والجنس.

د- تنوع وتشعب طرق الحصول على المعلومات، إذ يكفي أنه توجد على شبكة الوب مئات محركات البحث (Search Engines)، عدا الطرق الأخرى الموجودة في شبكة الإنترنت مثل Veronica وArchie وWAIS وغيرها.

ه- التعبير عن حاجة المستثمر التي أصبحت معقدة وصعبة النمذجة. ففي السابق كانت إمكانية الوصول إلى الوثائق عن طريق عناوينها أو بعض مواصفاتها (كالمؤلف أو تاريخ النشر أو دار

النشر) تعتبر إنجازاً جيداً ومقدّراً من قبل المستخدم. أما الآن، فلا يكتفي المستخدم بالوصول إلى هذه النتيجة، بل يطلب الوصول إلى الوثائق التي تحوي معلومات محددة ضمنها، بل وأكثر من ذلك، الوصول إلى أجزاء هذه الوثائق التي تحتوي المعلومات التي يريدها.

مما سبق يمكننا أن نقدّر حجم الصعوبات التي تواجهنا في عمليات البحث، فمشكلة الحصول على المعلومات المطلوبة بحد ذاتها هي مسألة محفوفة بالصعاب، وتزداد الأمور صعوبة عندما نحاول الحصول على هذه المعلومات بزمن معقول يجعل استخدام النظم مجدياً.

تركّز الحلول العملية، التي نجدها على شبكة الإنترنت في الوقت الحالي، على الحصول على نتائج البحث بزمن معقول مقابل ضعف في دقة نتائج هذا البحث.

1-4 التصنيف العريض لنظم البحث عن المعلومات

ذكرنا فيما سبق أن عمليات البحث التي تقوم بها أنظمة البحث متعددة، فمنها ما يؤمن عمليات بحث عن صفات خارجية للمعلومات، وأخرى تبحث ضمن المحتوى الدلالي للوثائق، والأخيرة تؤمن إمكان الانتقال المباشر إلى فضاء المعلومات والتجول فيها.

يقودنا هذا إلى اقتراح تقسيم أنظمة البحث (انظر الشكل 1-2) حسب طريقة البحث إلى الأنواع الثلاثة التالية:

1. نظم قواعد البيانات Databases

وهي تشكّل أيضاً البنية التحتية لأي نظام بحث عن المعلومات، وتسمح بالبحث عن المواصفات الخارجية للوثائق (العنوان، المؤلف، تاريخ النشر، ...).

2. نظم استرجاع المعلومات (Information Retrieval Systems)

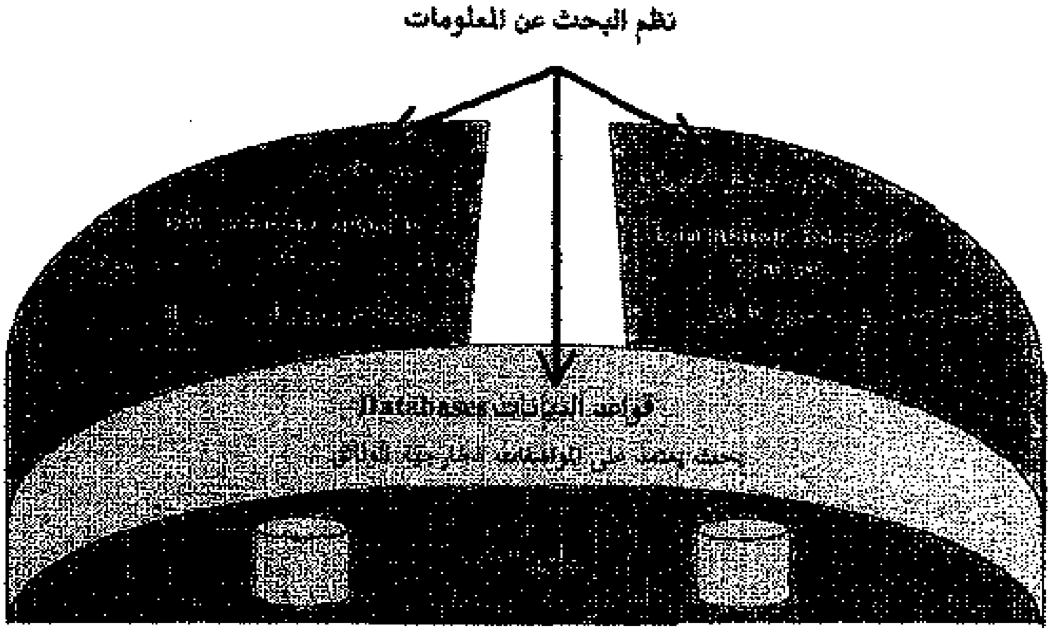
وفيها يجب التعرف على المحتوى الدلالي للوثائق (مضمون الوثائق)، ثم القيام بعمليات البحث حسب محتوى هذه الوثائق من المعلومات.

3. نظم التصفح (Navigational Systems)

وهي أكثر النظم حداثةً، وأسلوب البحث فيها هو الأكثر تفاعلاً مع المستخدم. يجري البحث فيها بواسطة التجول أو التصفح (Browsing) في فضاء الوثائق (Cyber Space) باستخدام الروابط (Links) والبنى المهيكلة (Structured units). فالمستخدم يجد نفسه في مواجهة مجموعة من الوثائق، ويزوده النظام بمجموعة من الارتباطات التي تتيح له التنقل بين الوثائق. يقوم المستخدم باختيار الارتباط الذي يريده حسب المعلومة التي يريد الوصول إليها.

4. النظم الهجينة (Hybrid Systems)

وهي النظم التي تسمح بمعظم عمليات البحث الآنفة الذكر (أي البحث عن مواصفات الوثائق، وعن محتواها، وتصفحها) بنسب تعقيد متفاوتة، وهذه هي حالة معظم محركات البحث (Search Engines) في الإنترنت.



الشكل (1-2): التصنيف العريض لنظم البحث عن المعلومات

1-5 خلاصة

لقد حددنا في هذا الفصل الإطار العام لمشكلة البحث عن المعلومات من وجهة نظرنا الشخصية (المتوافقة مع وجهات نظر العديد من الباحثين في هذا المضمار). وسنعرض في الفصل القادم الأنواع المختلفة لنظم البحث عن المعلومات ضمن هذا الإطار.

الفصل الثاني

نظم البحث عن المعلومات

2-1 مقدمة

نقدّم في هذا الفصل شرحاً نظرياً مبسطاً وسريعاً لكل نوع من أنواع نظم البحث عن المعلومات التي عرفّناها في الفصل الأول، والتي تدخل مفاهيمها في بناء أدوات البحث عن المعلومات في الإنترنت، ونركّز الاهتمام في نظم البحث غير المعروفة معرفةً جيدة من قبل القارئ الكريم.

نتعرض في المقطع الأول بعجالة للبحث عن المعلومات في قواعد البيانات (DataBases) التي تسمح للمستخدم بإجراء عمليات البحث حسب موصفَات (Attributes) المعلومات المخزنة سلفاً وفق بنية محددة.

أمّا في المقطع الثاني فسنشرح آليات البحث عن المعلومات في نظم استرجاع المعلومات (Information Retrieval Systems) التي تسمح للمستخدم بإجراء عمليات البحث على مضمون (محتوى) المعلومات (وربما هيكلتها ونوع الوسائط المؤلفة لها)، دون اهتمام كبير بموصّفات المعلومات المخزنة في النظام.

ننتقل في المقطع الثالث للتحدث عن آليات البحث عن المعلومات في نظم الوسائط الفائقة (Hypermedia Systems) التي تبتعد عن أسلوب سؤال المستخدم عن حاجته من المعلومات، وتسمح

له بالولوج المباشر إلى فضاء المعلومات، بحيث ينتقل (يتصفح) بين هذه المعلومات وفق الإمكانات المتاحة لهذه الأنظمة حتى الوصول إلى حاجته.

أخيرا، في المقطع الرابع نركز الاهتمام في البحث عن المعلومات في نظم تحتوي جميع (أو معظم) إمكانات البحث السابقة، وسنسمي هذه النظم النظم الهجينة (Hybrid Systems).

2-2 البحث عن المعلومات في نظم قواعد البيانات

تعالج نظم إدارة قواعد البيانات (Database Management Systems <DBMS>) ملفات من البيانات موصوفة بموصفات (Attributes) محدودة ومسبقة التعريف. فتوصف مثلا ذاتية الأشخاص العاملين في شركة ما بالموصفات التالية: اسم وشهرة الشخص – تاريخ ميلاده – عنوانه – وظيفته – راتبه – ... يخزن كل شخص في تسجيلة (Record) بحيث تكون لكل موصفة فيها قيمة ما (Value)، ومن ثم تعرف كل تسجيلة بقيم الموصفات فيها (مثلا يكون للتسجيلة الخاصة بي الشكل التالي: (1stName = "Ammar", 2ndName = "Kheirbek", DateofBirth = "6/10/1966", Address = "HIAST", Profession = "Researcher", Salary ="XXX", ...)).

تجري عملية البحث عن التسجيلات بتحديد قيم بعض الموصفات بواسطة طلب (Query) ليقوم النظام بعدئذ بإيجاد جميع التسجيلات التي تحقق هذه القيم، ويعتمد في ذلك على لغات استفسار بنيوية (Structure Query Language <SQL>).

إضافة إلى ذلك يستطيع نظام إدارة قواعد البيانات القيام بعمليات حسابية بسيطة مثل: حساب عدد التسجيلات التي تحقق قيم موصفات الطلب – القيمة الوسطية لموصفة ما – الانحراف المعياري...

حتى وقت قريب كانت هذه النظم وخصوصا العلائقية (Relational) منها هي نظم البحث الرائجة عن المعلومات. وقد كتب عنها الشيء الكثير، ولذلك لن ندخل في تفاصيلها ونترك ذلك للكتب المختصة. فيما يخص البحث عن المعلومات نجد أن الإمكانيات المتاحة للاستفسار في هذه النظم

مقصورة على البحث عن التسجيلات التي تحقق قيم موصفاتها خواص محدّدة (Strict)، وهذا لا يتناسب إلاّ قليلاً مع متطلبات المستخدمين. لكن التطورات الحديثة في هذا المجال مثل: نظم قواعد البيانات المعممة – نظم قواعد البيانات غرضية التوجه – نظم قواعد البيانات الموزعة – ... أتاحت إغناء عمليات البحث عن المعلومات لتشمل الوسائط المتعددة (Multimedia)، والمعلومات المهيكلة (Structured)، ... إضافةً إلى انفتاح هذه النظم على نظم استرجاع المعلومات والوسائط الفائقة وتطبيقات الإنترنت.

2-3 البحث عن المعلومات في نظم استرجاع المعلومات

سنعرّف في هذا المقطع بنظم استرجاع المعلومات (Information Retrieval Systems <IRS>)، ومن ثم نعرض المخطط العام لهذه النظم ووظائفها الأساسية، يلي ذلك عرضٌ سريع لأهم تقنيات استرجاع المعلومات والتي تستند إلى نماذج رياضية متينة ومتطورة.

2-3-1 تعريف نظم استرجاع المعلومات

هي النظم التي تسمح بتخزين وإدارة ومعالجة مجموعة من الوثائق (Documents) بطريقة تسمح للمستخدم باسترجاع الوثائق التي يتوافق محتواها مع حاجته من المعلومات.

يمكن أن تكون الوثيقة أي نوع من الوسائط (Media): نصيةً أو صوتيةً أو صوراً أو فيديو، كما يمكن أن تكون مهيكلةً (Structured) كما في حالة كتاب مكون من فصول، والفصول من أجزاء، والأجزاء من مقاطع، والمقاطع من فقرات. تشكّل مجموعة الوثائق قاعدة الوثائق (Corpus).

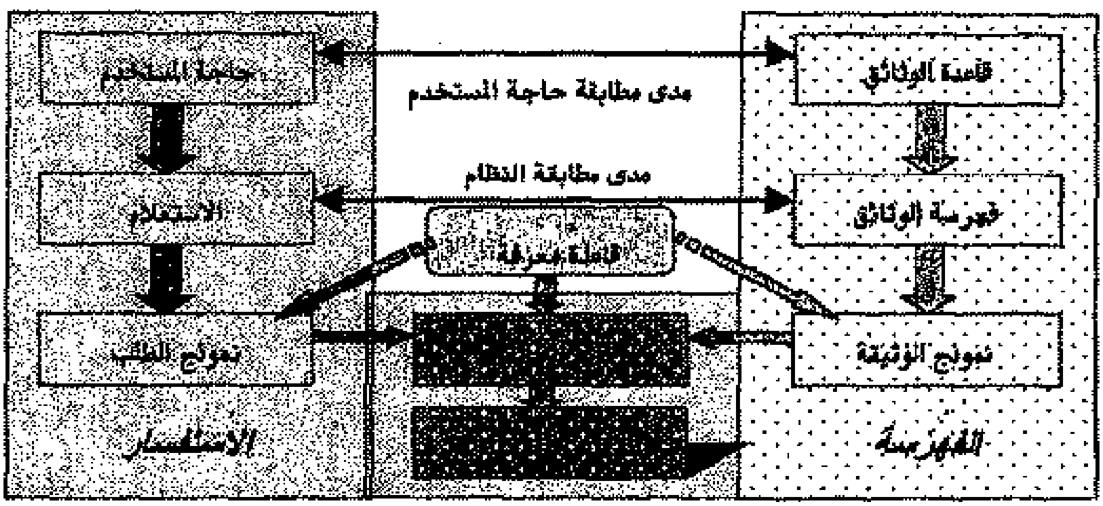
يرتبط بكل وثيقة مجموعة من المعارف (Knowledge) التي تعبر عن محتوى الوثيقة. وتشكّل المعارف المرتبطة بكل وثيقة ثلاثة أنواع من الموصفات:

- خارجية: عنوان، تاريخ، مؤلف، ... تسمى هذه الصفات خارجية لأنها تصف الوثيقة دون الاهتمام بالمادة التي تحويها.
- داخلية: وتعبر عن المحتوى الدلالي (Semantic Content) للوثيقة، أي موضوع الوثيقة. وتسمى داخلية لأنها تتعلق بمادة الوثيقة.
- بنيوية: البنية المنطقية (Logic Structure) للوثيقة، وتتعلق بكون الوثيقة مهيكلة أم لا، وبطبيعة الوسائط التي تحتويها.

يعبر المستخدم عن حاجته بطلب (Query) يحتوي تحديدا للموصفات (من أي نوع) المطلوبة، ويؤدي تنفيذه إلى اختيار الوثائق (أو الوثائق الجزئية) التي يعتبرها النظام موائمة له.

2-3-2 المخطط العام لنظم استرجاع المعلومات

يتألف نظام استرجاع المعلومات من ثلاثة أقسام أساسية هي (الشكل 2-1):



الشكل (2-1): البنية العامة لنظام استرجاع المعلومات

آ- الفهرسة

يجري في هذا القسم تحديد جميع موصفات الوثيقة (الداخلية والخارجية والبنيوية). فمن أجل المواصفات الخارجيَة، تُحدَّد معلومات مثل تاريخ النشر، والعنوان والمؤلف، ودار النشر، و... ومن أجل المواصفات البنيوية للوثيقة، تُحدَّد بنية هذه الوثيقة (فصول وأجزاء ومقاطع و...)، كما يُحدَّد نوع الوثيقة والوسائط التي تحويها (نصوص، صور، صوت، مختلطة...).

إن العملية الأكثر تعقيداً هي عملية التعرف على محتوى الوثيقة الدلالي، فهذه العملية تعني التعرّف على محتوى النص (في حال احتوت الوثيقة على نص فقط)، والتعرّض لمشاكل معالجة اللغات الطبيعية (تحديد بنية الجملة الفعلية كفعل وفاعل ومفعول به، أو بنية الجملة الاسمية، ومعالجة الحالات الإعرابية المختلفة وتعدّد المعاني للكلمة الواحدة، والالتباس والغموض في الجمل والتراكيب، والاصطلاحات، ...). إنّ حل هذه المشاكل هو أمر معقد جداً، ويزداد صعوبةً من أجل الوسائط الأخرى كالصورة أو الصوت، كما أنّه عمل متعلق باللغة الطبيعية المستخدمة، فحلها من أجل اللغة الإنكليزية مثلاً لا يعني إمكانية الحل من أجل اللغة العربية، وحتى الآن لا يزال قسم كبير من هذه المشاكل عالقاً فيما يتعلق باللغة العربية.

تعتمد معظم الطرق المتبعة في تحديد المضمون على تقنيات إحصائية ترتكز على إحصاء تواتر ورود بعض الكلمات المميزة في الوثيقة النصّية من أجل تحديد محتواها. بعد التعرّف على محتوى الوثيقة، يجب تمثيلها وفق نموذج معرفي معين (Knowledge Formalism) حتى نتمكن من القيام بعمليات حاسوبية عليها. نجد هنا مشاكل تتعلق بإمكانيات النموذج المعرفي المستخدم. إن المشاكل التي أشرنا إليها أكبر من أن تُذكر في هذه العجالة، ونكتفي بالإشارة إليها بسرعة لتكوّن لدى القارئ صورة عامة عنها. باختصار، يقوم النظام بمعالجة جميع الوثائق الموجودة في مجمع الوثائق، و"فهم محتواها"، ثم يمثّلها وفق نموذج معرفي خاص يمكّنه من القيام بعمليات وحسابات عليها فيما بعد.

ب- الاستفسار

وهو القسم الذي يتيح للمستخدم أن يتعامل مع النظام. يصوغ المستخدم المعلومات التي يريد البحث عنها على شكل طلب، مستخدماً لغة استعلام خاصة بالنظام، ويطلب من النظام البحث عن وثائق تحوي هذه المعلومات. يقوم النظام عند تلقي طلب البحث بالتعرف على طلب المستخدم وفهمه، ثم يصوغ هذا الطلب وفق نموذج معرفي خاص بالطلبات. يمكن لهذا النموذج المعرفي أن يُستخدم في الفهرسة أيضاً، كما يمكن له أن يكون مختلفاً حسب تصميم النموذج.

ج- تابع المطابقة

يقوم النظام في هذا القسم (وهو بنفس لون قسم الاستفسار لأنهما مرتبطان جداً أحدهما بالآخر) بمقارنة نموذج الوثيقة مع نموذج الطلب وفق تابع للمطابقة (Matching Function) معرّف سلفاً، وحساب درجة المواءمة بين الوثائق الأجوبة وطلب المستخدم، فيعطي درجات مواءمة عالية إذا كان محتوى الوثيقة يلبي الطلب الذي صاغه المستخدم، ثم يقوم بترتيب الوثائق التي تجيب على الطلب حسب درجة المواءمة المتناقصة. إن استخدام نموذج معرفي واحد لتمثيل الوثائق والطلبات يسهّل عمليةَ المطابقة. ويمكن استعمال تابع المطابقة للمقارنة بين أي وثيقتين في مجمّع الوثائق وذلك من أجل عمليات التصنيف والتكتيل (Clustering). سنعرض فيما بعد بعض نماذج المطابقة المستخدمة في أنظمة استرجاع المعلومات.

نلاحظ في الشكل (2-1) وجود جزء يسمى قاعدة المعرفة (Knowledge Base)، وهي تتكون من مجموعة من المعارف تسمح بالتعرف على علاقات مبطنة بين المفاهيم ضمن النص لا يمكن إيجادها آلياً، مثل المعرفة التي تخبرنا أن نبات البطاطا هو من الخضار، أو أن التفاح هو من الفاكهة. يمكن استخدام قاعدة المعرفة أيضاً من أجل دعم عمليات نمذجة الوثائق ونمذجة الطلبات، حيث تستخدم في كل من نموذج الوثيقة ونموذج الطلب وتابع المطابقة.

يوجد قياسان للمواءمة في هذه النظم:

* مواءمة المستخدم (User Relevance)

عندما يقترح النظام وثيقة كجواب على الطلب ويعتبرها المستخدم موائمة لحاجته.

* مواءمة النظام (System Relevance)

عندما يقترح النظام وثيقة كجواب على الطلب اعتماداً على تابع المطابقة الخاص به.

وكلما اقتربت مواءمة النظام من مواءمة المستخدم كلما زادت جودته. وتجدر الإشارة إلى أنّه لا يمكن لأي نظام أن يحقق مواءمة المستخدم، لأن ذلك يعني أن يكون لدى النظام نفس المعارف التي يمتلكها المستخدم (الإنسان) وطريقة معالجة المعارف نفسها، وهذا بعيد التحقيق لفترة طويلة قادمة !

يوجد أيضاً قياسان يحدّدان جودة النظام:

* الاستذكار (Recall) = عدد الوثائق الموائمة التي وجدها النظام / العدد الكلي للوثائق الموائمة في قاعدة الوثائق.

* الدقة (Precision) = عدد الوثائق الموائمة التي وجدها النظام / عدد الوثائق الكلي التي وجدها النظام.

2-3-3 تقنيات نظم استرجاع المعلومات

تستند تقنيات استرجاع المعلومات إلى نماذج رياضية متينة ومتطورة، وتقسم هذه التقنيات حسب نوعية تابع المطابقة إلى تقنيات مطابقة تامة و تقنيات مطابقة جزئية. عندما تكون المطابقة تامة فإن نموذج تابع المطابقة هو النموذج البولياني (Boolean Model). وعندما تكون المطابقة جزئية فإنها تقسم إلى نوعين حسب اعتبار الوثائق:

أ- فردية: حيث تُعتبر الوثيقة هي وحدة المقارنة ولا تُؤخذ بعين الاعتبار علاقتها مع الوثائق الأخرى ضمن مجمّع الوثائق. وتقسّم حسب الموصفات المستخدمة في المطابقة إلى قسمين: المطابقة المعتمدة على بنية الوثيقة والمطابقة المعتمدة على خواص ومحتوى الوثيقة. نجد في حال استخدام موصفات البنية نوعين من النماذج: البيانية (Graph) والمنطقية (Logic). و نجد في حالة استخدام موصفات المحتوى نماذج صورية (Formal Models) مثل: النماذج الشعاعية (Vecotorial Models)، والنماذج اللغوية (Linguistic Models)، والنماذج الاحتمالية (Probabilistic Models)، والنماذج المنطقية أو النظرية (Logic Models). سنذكر هذه النماذج بتفصيل أكبر فيما بعد.

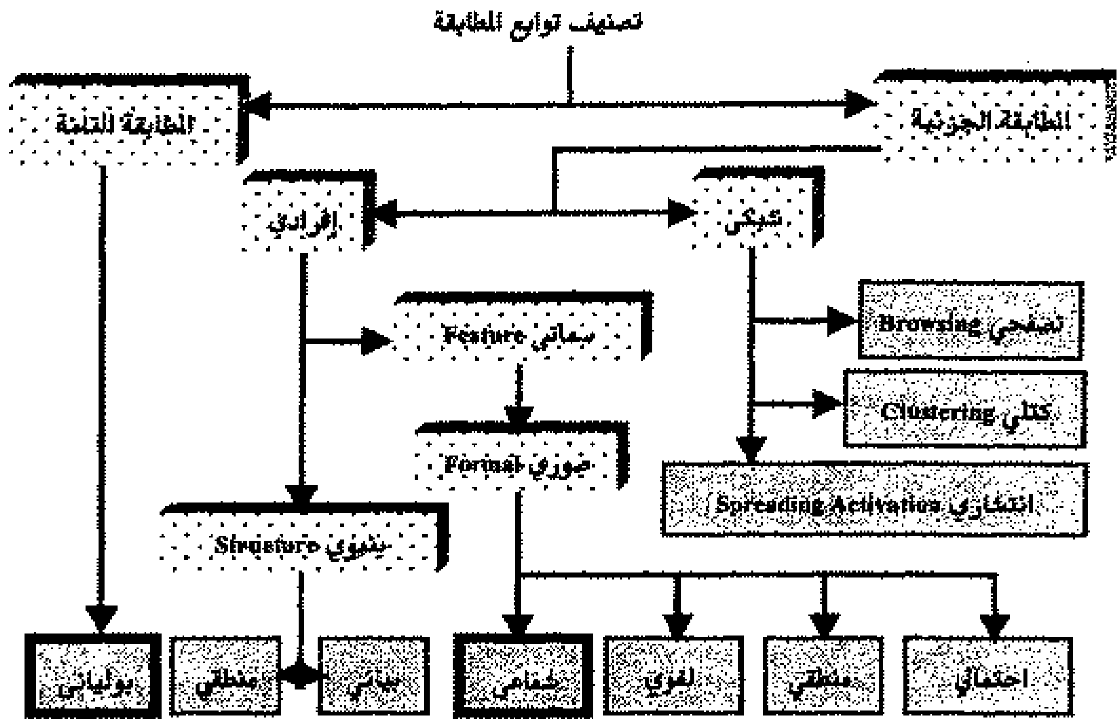
ب- شبكية: حيث تُؤخذ بعين الاعتبار العلاقات الممكنة بين الوثائق ضمن مجمّع الوثائق. ونجد هنا ثلاثة أنواع من النماذج: التصفحية (Browsing) والكتلية (Cluster-based) والانتشارية (Spread Activation).

يبين الشكل (2-2) التصنيف العام لتوابع المطابقة المختلفة.

2-3-3-1 المطابقة التامة والنموذج البولياني

تأتي المطابقة التامة من وجود الطلب في تمثيل الوثيقة تماماً. تُمثّل الوثيقة D في هذا النموذج كمجموعة من كلمات الترميز (أو الكلمات المفتاحية) التي تعبر عن محتوى الوثيقة (والتي حصلنا عليها نتيجة عملية الفهرسة) وترتبط بعملية و (And) حصراً. أما تمثيل الطلب Q فهو أيضاً مجموعة من كلمات الترميز تعبر عن حاجة المستخدم (والتي حصلنا عليها نتيجة عملية الاستفسار) ويمكن أن ترتبط بعمليات بوليانية متعدّدة مثل و، أو، نفي (And, Or, Not).

يعتمد تابع المطابقة على تقنية مطابقة النماذج (Pattern Matching)، أي أنه يبحث عن كلمات الترميز (الموجودة في الطلب) ضمن جميع وثائق قاعدة الوثائق، ويبني نتائجه على مبدأ الاقتضاء المؤكد $D \Rightarrow Q$ وهو بالتالي لا يعطي هذا التابع سوى إجابتين: مطابق أو غير مطابق.



الشكل (2–2): التصنيف العام لتوابع المطابقة

يعتبر تمثيل الوثائق في هذا النموذج بسيطا جدا، فهو يرتكز على مجموعة من كلمات الترميز (الكلمات المفتاحية) الموجودة في الوثيقة وتعبر عن محتواها الدلالي. من الصعب كتابة الطلبات في هذا النموذج بحيث تعبر جيدا عن حاجة المستخدم، ولكن من ناحية أخرى، يكون هذا النموذج فعالا من ناحية زمن الاستجابة. ولما كانت إجابة تابع المطابقة عن علاقة وثيقة ما بالطلب هي مطابقة أو غير مطابقة، فإننا لا نحصل على ترتيب للوثائق التي تجيب على الطلب بسبب الطبيعة الثنائية لتابع المطابقة، وهذا ما يؤدي إلى اعتباره قليل الدقة.

رغم بعض المساوئ التي ذكرناها، فإن هذا النموذج لم يزل مستخدما في معظم محركات البحث (Search Engines) على الإنترنت لأن السمة الأساسية التي تميزه عن غيره هي السرعة.

2-3-3-2 المطابقة الجزئية

يعتمد مبدأ تقنيات المطابقة الجزئية (Partial Matching Techniques) على وجود الطلب جزئيا في تمثيل الوثيقة. يستخدم تابع المطابقة عموما قياسات تقريبية لدرجة المطابقة، لذا تكون إجابته على طلب المستخدم من النمط: تجيب الوثيقة على الطلب بنسبة α التي يقوم النظام بحسابها.

يكمن الفارق الأساسي عن المطابقة التامة في أنه عند وجود مطابقة بين وثيقة والطلب، فليس من الضروري أن يكون الطلب محتوى في الوثيقة، وبالتالي توجد درجات للمطابقة تسمح بترتيب الوثائق التي تجيب على الطلب حسب مواءمتها له. يمكن الحصول على نفس فعالية المطابقة التامة إذا استخدمنا تقنيات خاصة (الملفات المعكوسة Inverted Files)، كما يؤمن هذا النموذج دقة أعلى بكثير من المطابقة التامة.

آ. النماذج الشعاعية

تمثل الوثائق والطلبات في النماذج الشعاعية (Vectorial Models) بأشعة في فراغ من t بعدا، حيث t هو عدد كلمات الترميز في قاعدة الوثائق كاملة. مسقط شعاع الوثيقة (أو الطلب) على محور كلمة الترميز ti هو تثقيل هذه الكلمة في الوثيقة (أو الطلب).

يحسب التثقيل (Weighting) اعتمادا على علاقة رياضية بين تكرار عبارة الترميز في الوثيقة وتكرارها في قاعدة الوثائق ككل (طريقة Tf/Idf). عندما يستطاع تمثيل الوثيقة والطلب كشعاعين، فإن حساب تابع المطابقة يؤول إلى حساب تجيب الزاوية θ بين شعاعي الطلب والوثيقة، أي أن قيمه ستكون محصورة بين 0 و1.

مشكلة هذا النموذج هي محدودية لغة الترميز وفقرها الدلالي واستقلال كلمات الترميز بعضها عن بعض.

ب. النماذج الاحتمالية

يقسم الطلب فضاء الوثائق في النماذج الاحتمالية (Probabilistic Models) إلى مجموعتين (وهنا يمكن اعتبارها مسألة تصنيف): مجموعة الوثائق الموائمة ومجموعة الوثائق غير الموائمة حسب القيم التي يأخذها تابع احتمالي معين. ويعتمد مبدأ ترتيب الوثائق الأجوبة على احتمال مواءمتها للطلب. ويؤدي استخدام طرق مختلفة لتقدير الاحتمالات إلى نماذج احتمالية مختلفة.

ميزتها عن النماذج الشعاعية هو إمكانية الأخذ بعين الاعتبار العلاقات بين كلمات الترميز. وغالبا ما تكون هذه النماذج صعبة البرمجة.

ج. النماذج اللغوية

تعتمد النماذج اللغوية (Linguistic Models) على تقنيات معالجة اللغات الطبيعية (Natural Language Processing) للوثائق والطلبات آليا. حلت هذه النماذج مشكلة محدودية لغة الترميز (التي أصبحت اللغة الطبيعية نفسها)، ولكنها وقعت في مشاكل أكثر تعقيدا أهمها غموض التعبير والاستفسار في اللغة الطبيعية، وإعطاء الإيحاء للمستخدم بأن النظم التي تعتمدها هي نظم مفتوحة ويمكن أن تقبل أي طلب.

يختلف تعقيد المعالجة وتابع المطابقة حسب درجة تعقيد التمثيل الدلالي (ارتباطات معرفية – شبكات دلالية – بيانات معرفية – ...).

الميزة الأساسية في هذه النماذج هي أن المستخدم يستطيع التعبير عن حاجته بلغة طبيعية دون معرفة النظام. تعتبر هذه النماذج عموما معقدة جدا ويفضل تطبيقها في مجالات محددة تماما.

د. النماذج المنطقية أو النظرية

يمكن اعتبار النماذج السابقة نماذج عملية (Operational Models) أو إحصائية. وقد وصل أداء نظم استرجاع المعلومات المعتمدة على النماذج الإحصائية إلى حد الإشباع النظري. فكان لا بد من إيجاد

نظرية جديدة وعامة من أجل بناء جيل جديد من نظم استرجاع المعلومات. تتجلى هذه العمومية في استقلال النموذج عن التمثيل الدلالي للوثائق والطلبات.

المبدأ الأساسي في النماذج المنطقية أو النظرية (Theoretical or logic Models) هو الاقتضاء غير المؤكد (Uncertain Implication):

الوثيقة D تلائم الطلب Q إذا كانت تحقق الاقتضاء غير المؤكد D → Q، وتابع المطابقة P هو الذي يقيس درجة هذا الاقتضاء.

يعتمد الاقتضاء غير المؤكد على المبدأ العام للشك (Incertitude):

يحدد قياس الشك في X→Y بالنسبة إلى مجموعة معطاة من المعارف بالتوسع الأصغري لهذه المجموعة من أجل تثبيت الحقيقة لـ X→Y. حتى نتمكن من تثبيت الحقيقة لـ D→Q من أجل قاعدة معرفة K، نقوم بتغيير D أو Q أو K وهذا ما يسمى بالمبادئ الثلاثة للشك.

يقيس تابع المطابقة عدديا درجة التغيير التي نجريها حسب أحد المبادئ الثلاثة للشك.

اختيار المنطق المستخدم مهم جدا من أجل أداء النظام. وتوجد مجموعة من التجارب المعروفة التي تستخدم النموذج المنطقي هي: المنطق البولياني – المنطق الشكلي (Modal Logic) – المنطق الشكلي الغائم (Fuzzy Modal Logic) – المنطق الغائم (Fuzzy Logic) – المنطق المصطلحاتي (Terminological Logic).

وقد أثبتت البحوث الحديثة إمكانية استيعاب بنية الوسائط الفائقة في النموذج من خلال استعمال الاقتضاء المعكوس.

هـ. النماذج البنيوية

يمدد تمثيل المعارف في النماذج البنيوية (Structure-based Models) ليشمل البنية المنطقية للوثائق وذلك باستعمال منطق صوري (Formal Logic) (المنطق الفرداتي مثلا) أو باستعمال البيانات

(البيانات المرفية Conceptual Graphs مثلا). وتتحول عملية المطابقة إلى عملية استدراكية غير مؤكدة (Uncertain Inference).

يمكن حديثا استخدام لغات لتمثيل البنية المنطقية للوثائق مثل:

SGML: Structured Generalized Markup Language

HTML: Hypertext Markup Language

و. النماذج الشبكية

النماذج الشبكية (Network Models) هي النماذج التي تأخذ بعين الاعتبار العلاقات الممكنة بين الوثائق كنماذج التفعيل الانتشاري (Spread Activation) والتصفح (Browsing). تفترض نماذج التفعيل الانتشاري وجود شبكة من مفاهيم الترميز مرتبطة مع الوثائق ويمكن للمستخدم التجول فيها وعند وصوله إلى المفهوم المطلوب يستطيع استكشاف الوثائق المرتبطة به. بمعنى آخر يمكن تصور ذلك كشبكتي وسائط فائقة إحداهما للمفاهيم (Hyperconcept) والثانية للوثائق (Hyperbase) مع إمكانية التنقل بينهما.

أما نماذج التصفح فهي مستقاة من مجال الوسائط الفائقة (سندرسها لاحقا)، وهي تسمح للمستخدم بالتجول ضمن قاعدة الوثائق باتباع مسارات خاصة حسب المعلومة التي يريد الوصول إليها.

النوع الأخير من النماذج الشبكية هي النماذج الكتلية (Cluster-Based Models)، التي تستخدم خوارزميات التكتل (Clustering) في تجميع الوثائق المتشابهة، ثم تحديد ممثلة عن كل كتلة. وعندما يطلب المستخدم من النظام البحث عن الوثائق، يكتفي النظام بمقارنة الطلب مع ممثلات الكتل فقط مما يختصر المقارنات مع جميع الوثائق، فإذا كانت ممثلة الكتلة جوابا للطلب يلجأ النظام إلى المقارنة مع الوثائق الموجودة في هذه الكتلة.

2-3-4 عرض لبعض نظم استرجاع المعلومات

نعرض في هذه الفقرة بعض النماذج المخبرية (Prototypes) الأكاديمية، بهدف إعطاء القارئ فكرة عن كيفية تعريف نماذج نظم استرجاع المعلومات مع ما ينطوي عليه ذلك من صعوبات مختلفة، وليس بهدف تعريفه على تطبيقات تجارية معروفة يستطيع استخدامها. إذ تبقى محركات البحث في الإنترنت تطبيقات مهمة لنظم استرجاع المعلومات رغم اقتصارها على مجموعة ضئيلة من إمكانيات هذه النظم.

2-3-4-1 النظام IOTA

طور النظام IOTA في المخبر LGI في جامعة جوزيف فورييه في فرنسا، ويهدف إلى إدارة مجمع من الوثائق المهيكلة، وتنفيذ طلبات الاستعلام التي تجري عليها.

تعاملت أنظمة استرجاع المعلومات قبل النظام IOTA مع الوثائق على أنها ذات بنية عنصرية (Atomic) لا تتجزأ، أما النظام IOTA فقد تمكن من مكاملة مفهوم الوثائق المهيكلة ضمن نظام استرجاع المعلومات، سواء على مستوى الفهرسة أو على مستوى معالجة طلبات الاستعلام.

الفهرسة في النظام IOTA آلية، وتستند على تقنيات لغوية مع مراعاة بنية الوثيقة المنطقية (فصول، مقاطع، فقرات ...)، والاهتمام بالعناصر النصية المميزة الموجودة ضمن الوثيقة مثل المقدمة والخاتمة وعناوين الفقرات.

يحلل النظام النص ثم يستخرج مجموعة الكلمات المميزة (المفتاحية) فيه. بعد ذلك، يستخدم توابع إحصائية لتمثيل العلاقة المتبادلة بين كل من هذه المفردات والوحدات البنيوية التي تتألف منها الوثيقة (الفصول، المقاطع، الفقرات ...)، فيعرف مفهومين جديدين هما: تمثيل الكلمة للوحدة، وتمثيل الوحدة للكلمة. تمثيل الوحدة للكلمة هي قيمة بين الصفر والواحد تساوي عدد مرات ورود الكلمة ضمن تلك الوحدة مقسوما على عدد مرات ورود كلمات الفهرسة جميعا ضمن تلك الوحدة. بمعنى آخر، كلما ازداد عدد مرات ورود كلمة ضمن وحدة، كلما كان تمثيل هذه الوحدة كبيرا. أما

تمثيل الكلمة للوحدة فهي قيمة بين الصفر والواحد أيضا، وتساوي عدد مرات ورود الكلمة ضمن الوحدة مقسوما على عدد مرات ورود الكلمة ضمن جميع الوحدات في مجمع الوثائق. بمعنى آخر، إذا تكررت إحدى الكلمات تكرارا كبيرا في وحدة ما بالتحديد، فإن تمثيل هذه الكلمة لتلك الوحدة سيكون كبيرا. يمكن التعبير عن التمثيل المتبادل بين كلمة ووحدة بالتوسط الحسابي بين تمثيل الوحدة للكلمة وتمثيل الكلمة للوحدة.

تبدأ عملية الفهرسة من الوحدات الأوراق في بنية الوثيقة الهرمية. أما الوحدات في المستويات الأعلى، فتفهرس بناء على نتائج فهرسة الوحدات الأدنى. فتمثيل الكلمة للوحدة هو نسبة عدد مرات تكرار الكلمة ضمن جميع الأبناء إلى عدد مرات ورود كلمات الفهرسة جميعا ضمن جميع الوحدات الأبناء. أما تمثيل الوحدة للكلمة فهو مجموع تمثيل الوحدات الأبناء لتلك الكلمة.

يعتمد تابع المطابقة على النموذج الشعاعي، ويكتب الطلب كمجموعة من الكلمات المفتاحية من خلال واجهة تخاطبية تسمح أيضا بتحديد الوحدات البنيوية المطلوبة.

2-3-4-2 النظام RIME

طور النظام RIME في المخبر LGI في جامعة جوزيف فورييه في فرنسا، وهو يعالج مجمعا من الصور الشعاعية للمرضى في المستشفيات.

يرفق بكل صورة شعاعية شرح نصي عن هذه الصورة، يعبر عما تمثله الصورة، أي المحتوى الدلالي للصورة. يستخدم النظام RIME هذا الشرح من أجل فهرسة الصور الشعاعية، وتحديد محتواها. إذن هو لا يقوم بمعالجة الصورة نفسها، وإنما يحدد محتواها من خلال معالجة الشرح المرفق بها.

يستخدم النظام RIME معجما ومكنزا (Thesaurus). يحتوي المعجم على الكلمات المختلفة المعبرة عن الأشكال الممكن ظهورها ضمن الصور الشعاعية، ويستخدم لتفسير الصور الشعاعية. أما المكنز فيحتوي على مجموعة من المفاهيم الأولية للمجال الذي يتعامل معه النظام، وهو المجال الطبي الذي تغطيه هذه الصور، إضافة إلى العلاقات الدلالية بين هذه المفاهيم مثل علاقات التخصيص والتعميم،

كما في قولنا المفهوم A أعم من المفهوم B، أو علاقات التضمين كما في قولنا المفهوم A جزء من المفهوم B. يستخدم النظام RIME نموذج الارتباط المفاهيمي (Conceptual Dependency) لتمثيل المعارف الدلالية الموجودة فيه.

لفهرسة صورة شعاعية، تفهرس المعلومات النصية وتحول إلى بنى مفهومية تعبر عن المحتوى الدلالي للصورة الشعاعية، وممثلة وفق النموذج المعرفي المستخدم، أي الارتباط المفاهيمي.

يعالج النظام RIME النصوص بطريقة معقدة، وباستخدام خوارزميات معالجة اللغات الطبيعية، ويستخدم النموذج المنطقي في عملية المطابقة، ويسمح باستخدام لغة استفسارات طبيعية لصياغة الطلبات.

2-3-4-3 النظام EXPRIM

طور النظام EXPRIM في المخبر CRIN في جامعة نانسي في فرنسا، وهو أول نظام استرجاع معلومات يطبق على الصور. يتعامل النظام EXPRIM مع قاعدة معطيات من الصور، وقاعدة معطيات أخرى تحوي شرحا عن محتوى هذه الصور.

يتيح النظام EXPRIM القدرة على استخدام التغذية الراجعة في معالجة طلبات الاستعلام. فالنظام يسمح للمستخدم بأن يحدد، من مجموعة الصور التي يعرضها للمستخدم، الصور التي يجدها مطابقة لحاجته، والصور غير المطابقة، ثم يعيد النظام البحث من جديد آخذا بعين الاعتبار ما حدده المستخدم. ويمكن للمستخدم تكرار هذه العملية حتى يحصل على نتيجة مناسبة له. وتسمى هذه العملية في مفردات الذكاء الصنعي (Artificial Intelligence) التعلم (Learning).

يستخدم النظام EXPRIM مكنزا يفيد على السواء المستخدم في صياغة الطلبات، والنظام في معالجتها. كما يستخدم النموذج الشعاعي لتمثيل الوثائق والطلبات وكنموذج للمطابقة.

2-4 البحث عن المعلومات في نظم الوسائط الفائقة

ننتقل الآن إلى استعراض نمط آخر من نظم البحث عن المعلومات والذي لاقى منذ انتشاره الفعلي في أواخر الثمانينيات نجاحا منقطع النظير.

2-4-1 مقدمة

ظهرت فكرة النصوص الفائقة (Hypertext) في العام 1945، عندما نشر الصحفي Vannvar Bush، مقالته الشهيرة "كما يمكن أن نفكر"، التي بين فيها أن الدماغ البشري يعمل بطريقة تجميعية (Associative) وفق مبدأ ترابط الأفكار. وتجب محاكاة طريقة التفكير هذه بتقديم المعلومات النصية في أجزاء يمكن أن تكون مستقلة، على أن ترتبط الأجزاء المتوافقة في المضمون بعضها ببعض بواسطة روابط يلجأ إليها القارئ عند الحاجة. وقد اقترح Bush بناء النظام Memex، الذي يعتمد بنية غير خطية لتمثيل النصوص، مما يمكن المستخدم من قراءة هذه الوثائق بطريقة تجميعية، أي بالانتقال بين أجزاء النصوص حسب مواضيعها، وليس حسب الترتيب الخطي الذي تفرضه. ظل هذا النظام على الورق نظرا لضعف الإمكانات التقنية في ذلك العصر، إلا أن Vannvar Bush تنبأ بقدوم اليوم الذي يسمح فيه التطور التقني بتحقيق هذا النظام.

وكان لا بد من الانتظار حتى العام 1963 كي يأخذ Douglas Engelbart (مخترع الفأرة Mouse) أفكار Vannvar Bush ويطبقها في نظامه NLS (oN Line System) الذي يعتبر نصف نظام نصوص فائقة. لقد أراد Engiebart البرهان على أن استخدام الحاسوب في عمليات تنظيم ومعالجة المعلومات بآلية النصوص الفائقة يساعد على تطوير قدرة الاستيعاب لدى المستخدم.

أما مصطلح الـ Hypertext فقد اخترعه Ted Nelson في عام 1967 عندما صمم نظامه Xanadu الذي سمح للكثير من المستخدمين بأن يصلوا إلى معلومات مختلفة صوتية ونصية وبصرية.

ومع تطور وسائط الاتصال، وظهور الحواسيب المنزلية في بداية الثمانينيات، ظهرت أفكار متعددة من أجل دمج الصوت والصورة ضمن برامج حاسوبية بحيث تزيد من تفاعل المستخدم معها. وهكذا غزت ألعاب الفيديو المنازل حاملة التقنيات الجديدة، لكن المصطلحات مثل وثائق فائقة ووسائط فائقة ظلت محصورة في أشخاص على اطلاع على أعمال Theodor Nelson.

وفي عام 1987 بدأت مرحلة جديدة من تاريخ الوسائط الفائقة، فقد حدث أمران مهمان أحدثا تبدلا كبيرا في طريقة التعامل مع الوسائط الفائقة. الحدث الأول كان إصدار برنامج الوسائط الفائقة HyperCard الذي أنتجته شركة Apple، أما الحدث الثاني فكان انعقاد المؤتمر العالمي الأول للوسائط الفائقة برعاية ACM (Association of Computer Machinery) الذي نجح في دفع الباحثين نحو الاقتناع بأهمية الوسائط الفائقة، وساعد على ذلك التطور في تقنيات واجهات التخاطب مع المستخدم وانخفاض كلفتها، والتطور في وسائط التخزين، وظهور تطبيقات تتطلب إدارة مرنة وسهلة لحجوم كبيرة من المعلومات.

وفي عام 1988 اجتمعت مجموعة من المصممين في فندق ديكستر في مدينة نيو هامبشير في الولايات المتحدة من أجل تعريف وتحديد مصطلحات وطرق تصميم الوسائط الفائقة، وكانت النتيجة ظهور نموذج لنظام وسائط فائقة هو نموذج Dexter، وفيه وصفت أجزاء النظام الأساسية بطريقة غير صورية، كما احتوى على قواعد تقييم من أجل مقارنة النظم الموجودة في ذلك الوقت بعضها ببعض.

كانت الميزة الأكثر أهمية في نموذج ديكستر هي الفصل المنطقي بين محتوى الوثائق وبنية هذه الوثائق التي تفرضها الارتباطات الموجودة بين الوثائق. قام المصممون بتحقيق ذلك بأن أوجدوا طريقة لتوليد الارتباطات بطريقة تسمح للعديد من البنى (شبكات الارتباطات) بالتواجد معا.

عرف مصطلح الوسائط الفائقة (Hypermedia) على أنه المكاملة بين النص والصوت والرسوم والصور الساكنة والصور المتحركة ضمن نظام حاسوبي واحد. أخذ هذا المفهوم أهمية كبيرة بعد ظهور إمكانات تحويل الوسائط المتعددة إلى نسخ رقمية، فالنسخة الرقمية يمكن استرجاعها من أي مكان نريد، وهذا

ما يسمح بتجزئة الوسائط المختلفة ثم عرضها بالشكل الذي نريد، الأمر الذي يفتح آفاقا واسعة جدا في الاتصال والتخاطب.

وقد عانت نظم الوسائط الفائقة من عدم وجود ضوابط في تصميمها أو حتى من أجل مقارنتها، فكان القاسم المشترك الوحيد الممكن استخدامه للمقارنة بين مختلف الأنظمة هو تأمين تخاطب عال مع المستخدم. وضمن التنوع والاختلاف القائم، كان لا بد من إيجاد ضوابط من أجل ضمان التكامل والتوافق بين المنتجات على اختلاف الأنظمة التي تعمل عليها (اختلاف فيزيائي أو برمجي)، وهذا يتضمن وضع قواعد للتحويل المرئي للصور والنصوص والصوت، وقواعد من أجل الضغط (ضغط الصور والصوت) وقواعد للحفظ والتخزين. لكن وضع هذه الضوابط ليس كافيا، فالشركات لن تلتزم بها إن وجدت أنها تعيق إنجاز منتجات لها فائدة في الميدان العملي، مثل ميداني التسالي والتعليم بمساندة الحاسوب.

حدثت النقلة النوعية الثانية في تاريخ الوسائط الفائقة عام 1993 عندما وضع متصفح (Browser) Mosaic قيد الاستخدام. فمنذ أواخر الثمانينيات وأوائل التسعينيات بدأ الحديث عن شبكة الإنترنت ومعاييرها القياسية، ومفاهيم الشبكة العنكبوتية العالمية (World Wide Web <WWW>) أو الوب كنظام وسائط فائقة كوني يجمع مصادر المعلومات من جميع الحواسيب المتصلة على الإنترنت، وفق بروتوكول HTTP ولغة تأشير النصوص الفائقة HTML. لكن كان يجب انتظار أول متصفح قادر على تفسير وإظهار النصوص الفائقة. منذ 1993 وحتى الآن تعددت المتصفحات والخدمات التي يقدمها الوب حتى أصبح الوب النظام الأكثر استخداما من أجل البحث عن المعلومات، ونشرها، والتسويق، والتجارة الإلكترونية، ...

ربما تغيب الآن عن بال الكثيرين مفاهيم الوسائط الفائقة التي أدت إلى انتشار استخدام الوب على هذا النحو الواسع. لذلك وجدنا أنه سيكون من المفيد في هذا الكتاب أن نعود قليلا إلى الوراء لتوضيح وشرح مفاهيم الوسائط الفائقة، من أجل استيعاب أبعاد هذه النظم بهدف استخدامها الأمثل، وخصوصا استخدام الوب الذي يحتوي معظم المعلومات التي يمكن تخيلها. في هذا المقطع سنتعرف بمفاهيم الوسائط الفائقة وكيفية البحث عن المعلومات فيها.

2-4-2 تعريف الوسائط الفائقة

يمكن تعريف النصوص الفائقة (Hypertext) بأنها تمثيل غير خطي (Non Linear) لمعلومات نصية، على شكل بيان (Graph) من العقد (Nodes) المرتبطة (Linked) بعضها ببعض. تسمح هذه الروابط (Links) بالتصفح/التجول/الإبحار (Navigation / Browsing) ضمن العقد بطرق مختلفة، مما يعني تفاعلا كبيرا بين المستخدم ونظام النصوص الفائقة. يخضع اختيار طرق التصفح لمبدأ تجميع الأفكار والمعلومات، والمحاكمة بالمشابهة.

الوسائط الفائقة (Hypermedia) هي تعميم للنصوص الفائقة، إذ تكون فيها المعلومات من أي نوع من الوسائط (صوت، صورة، نص، فيديو،...)، بعبارة أخرى يمكن القول أن الوسائط الفائقة هي المزاوجة بين النصوص الفائقة والوسائط المتعددة (Multimedia).

تتصف معظم أنظمة الوسائط الفائقة بأنها توزع المعلومات ضمن عقد، ثم تربط بينها بروابط تسمح للمستخدم بالتجول من عقدة إلى أخرى. كما تؤمن هذه الأنظمة مجموعة من الأدوات من أجل المعالجة الساكنة والديناميكية للعقد والروابط، وتخزين المعلومات، وكتابة طلبات بحث بسيطة (بحث عن كلمة) والمساعدة في التصفح، و...

ومن ناحية التخاطبية مع المستخدم، تظهر العقدة للمستخدم كنافذة على الشاشة تحوي معلومات، أما الروابط فتظهر كمناطق حساسة (Anchors) على شكل أزرار (Buttons) أو ارتباطات تشعبية (Hyperlinks) ضمن النافذة. أما عملية التصفح والانتقال بين العقد فتتم بسهولة من خلال ضغط زر الفأرة على المنطقة الحساسة، فيقوم النظام بإظهار العقدة المطلوبة على نفس النافذة أو على نافذة أخرى. يظهر الشكل (2-3) مخططا عاما لنظام وسائط فائقة يمكن أن يحتوي صوتا وصورة وفيديو إضافة للنص.

2-4-3 بنية نظم الوسائط الفائقة

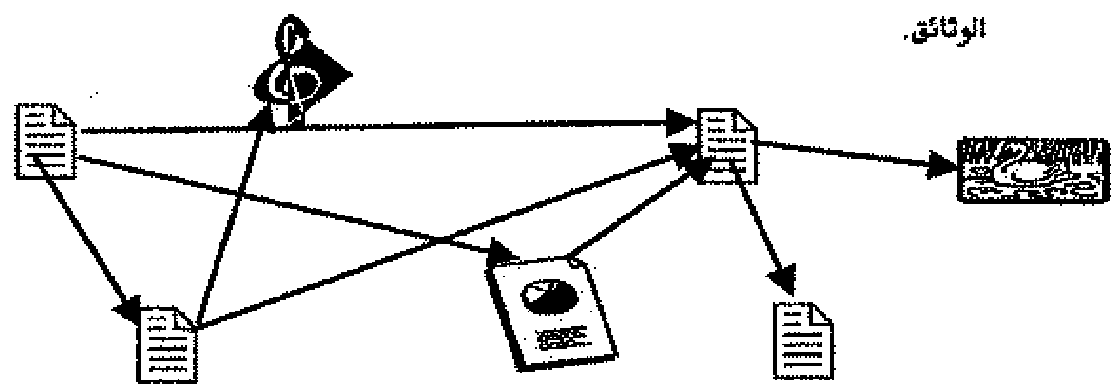
العناصر الأساسية في بنية نظام وسائط فائقة هي: العقد، والروابط، والمناطق الحساسة ضمن الوثيقة، وهذه الأخيرة هي بمثابة عنوان ينطلق منه ارتباط أو يصل إليه.

2-4-3-1 العقد

تشكل العقد في نظام وسائط فائقة أحد أهم عناصر بنيته، وتقسم هذه العقد إلى نوعين هما:

أ- عقد المعلومات (Information Nodes): وهي الوحدات الصغرى التي جزأنا إليها الوثيقة، وتحوي معلومات من نمط وحيد أو من عدة أنماط: نص، صورة، صوت،... ففي وثيقة متعددة الوسائط، توجد عقد تحوي النص (عقد نصية)، كما توجد عقد أخرى تحوي الصور (عقد صور)، وأخرى تحوي مقاطع الصوت (عقد صوت)، وعقد مختلطة تحوي مجموعة من العقد النصية وعقد الصور وعقد الصوت.

ب- العقد المحسوبة (Calculated Nodes): وهي عقد يجري حسابها آليا، أو شبه آليا، انطلاقا من عقد المعلومات، وهي تفيد في هيكلة وتنظيم قاعدة المعلومات. تحسب هذه العقد انطلاقا المتعلقة بالمعلومات عن بنية الوثيقة، أو من معلومات إحصائية عن العقد ضمن قاعدة الوثائق.



الشكل (2-3): الوسائط الفائقة والروابط بين الوسائط المختلفة

2-4-3-2 الروابط

إن التنوع الكبير في أنواع العقد التي يمكن لنظام وسائط فائقة أن يحتويها هو الذي يعطي هذا النظام المرونة الكبيرة، كما أن هذه المرونة تزداد بوجود التنوع الكبير في الروابط بين هذه العقد. تصنف هذه الروابط في عدد من المجموعات هي:

أ- **الروابط المحسوبة (Calculated Links)**: وهي مجموعة الروابط التي تولد عند حاجة المستخدم، وتسمى أيضا الروابط الديناميكية (Dynamic Links)، لأنها لا توجد فعليا إلا عند طلب المستخدم لها. تحسب هذه الارتباطات أثناء تصفح المستخدم، فعندما يصل إلى عقدة معينة، يستطيع الطلب من النظام أن يحدد له مجموعة العقد الممكن له الانتقال إليها والتي تحتوي معلومات قد يحتاجها، عندها يقوم النظام بمقارنة العقد ضمن مجمع الوثائق مع العقدة الحالية، ثم يحدد العقد المناسبة للمستخدم. حتى تتحقق الفائدة من هذه الروابط، يجب أن يجري حسابها بسرعة، وفي أحيان كثيرة يمكن الاستعانة بقاعدة بيانات من أجل تسريع عملية إيجاد هذه الروابط، حيث يقوم النظام بحساب الروابط بين العقد مسبقا، ثم تخزينها ضمن قاعدة البيانات. وعند حاجة المستخدم للارتباطات الديناميكية، يقوم النظام بالبحث ضمن قاعدة المعطيات عن الروابط المناسبة والتي قام بحسابها مسبقا.

ب- **الروابط البنيوية (Structured Links)**: وتبنى حسب بنية الوثيقة، فتربط بين عقدة أب والعقد الأبناء لها (وفق البنية المنطقية للوثيقة والمفروضة من قبل الكاتب). من الناحية الدلالية، تربط هذه الروابط بين العقد حسب درجة التفصيل في شرح المعلومات التي تحويها، فتنتقل المستخدم من معلومات عامة إلى معلومات أكثر تفصيلا وتخصصا.

ج- **روابط الفهرسة (Index Links)**: وهي روابط مشابهة لتلك الموجودة في النسخ الورقية، ففي نهاية بعض الكتب، توجد مجموعة من الكلمات الهامة مع أرقام الصفحات التي

تحويها. إن عمل هذه الروابط مشابه لعمل مجموعة الصفحات المذكورة، فهي تتيح للمستخدم إمكانية الانتقال بين العقد التي تحوي هذه الكلمات. يمكن لهذه الروابط أن تستخدم اسم المؤلف أو اسم الكتاب في عمليات الربط بين العقد. تبنى هذه الروابط يدوياً، أو آلياً، فيقوم النظام بفحص محتوى العقدة والحصول على المعلومات التي يحتاجها من أجل إنشاء الروابط.

د- **الروابط العكسية (Inverse Links)**: إن إنشاء روابط باتجاه وحيد يجعل قاعدة الوثائق غير متناظرة، لذا يفضل أن يتم إنشاء الارتباط العكسي عند إنشاء أي ارتباط مهما كان نوعه.

هـ- **الروابط ذات القيم (Weighted Links)**: تكمن سيئة البناء الآلي للروابط في كونه يقود إلى إنشاء عدد كبير منها، لذا من الفضل وجود قيم، تضاف إلى الروابط الناتجة، وتدل على درجة مطابقة هذه الروابط بين العقد. وترتب هذه الروابط بحيث يمكن للمستخدم التصفح حسب الروابط ذات القيم العالية، مما يضمن له وصولاً سريعاً إلى العقد المطلوبة.

و- **الروابط المنمطة (Typed Links)**: وهي روابط لها أنماط مسبقة التعريف، مثل روابط التعميم والتخصيص (Is more general, Is more special) وروابط الانتماء أو الاحتواء (Belongs to, Is part of) وروابط التوضيح (Illustrated by) وغيرها.

ز- **روابط الاقتضاء المنطقي (Implication Links)**: وهي روابط بين العقد التي ترتبط بعلاقات سببية.

ح- **روابط تشغيل (Function Links)**: وهي تتمثل في الأزرار التي تظهر على شاشة المستخدم، والتي تسمح بتشغيل برامج مرتبطة بها عند الضغط عليها باستخدام الفأرة،

ويمكن لهذه البرامج على سبيل المثال أن تقوم بالوصول إلى قاعدة معطيات لجلب معلومات خاصة بالعقدة الحالية وعرضها على المستخدم.

ط- روابط التزامن (Synchronization Links): وهي روابط يتوقف وجودها على الحاجة، ويمكن أن نحتاجها إذا أردنا إيجاد تزامن بين عقدتين. فمثلا إذا أردنا عرض عقدة فيديو وعقدة صوت (مرافقة لها)، فيجب أن يوجد تزامن بين العقدتين عند العرض، وهنا يمكن استخدام رابطة التزامن بينهما.

2-4-4 بعض إشكاليات نظم الوسائط الفائقة

يعتبر مجال الوسائط الفائقة من العلوم الحديثة نسبيا (عمليا منذ عام 1987)، وهو يتطور بسرعة كبيرة بسبب الإنترنت، وبالتالي مازال حقلا خصبا للبحث العلمي. لذلك يوجد فيه الكثير من الإشكاليات التي تمثل تحديات للعاملين في هذا الحقل، من هذه الإشكاليات نذكر ما يلي:

2-4-4-1 إدارة نظم الوسائط الفائقة

إن كثرة العقد والروابط في نظام وسائط فائقة تجعل من المستحيل عمليا القيام بصيانة النظام والتحقق من تماسك قاعدة الوثائق يدويا، لذا يجب على النظام أن يؤمن مجموعة من الأدوات التي تمكن من القيام بعمليات التحقق والصيانة عند الحاجة.

لنأخذ الحالة التي يطلب فيها إجراء تعديل على قاعدة الوثائق. مثلا نريد إضافة رابطة جديدة إليها، وهي حالة سهلة. في هذه الحالة يجب تحديد العقدة التي ستنطلق منها الرابطة الجديدة، ثم تحديد العقد التي ستشكل الطرف الآخر من الرابطة ضمن مجمع الوثائق. بعد ذلك، يضاف هذا الرابط إلى مجموعة الروابط الموجودة في النظام، وتحدد مناطق حساسة ضمن العقد المعنية من أجل هذه الرابطة الجديدة.

تصبح العملية أكثر تعقيدا عند حذف رابطة من قاعدة الوثائق، فحذف هذه الرابطة قد يؤدي إلى اختلال قاعدة الوثائق، فقد تنتج عقد ضمن قاعدة الوثائق لا يمكن الوصول إليها عبر أي طريق تصفح ممكن، أو قد يؤدي هذا الحذف إلى حذف معلومات دلالية أكثر من تلك التي نريد حذفها فعلا، وهذا يعود إلى تعقيد شبكة العلاقات ضمن قاعدة الوثائق، فلا يمكن التأكد من نتائج التعديل الذي نقوم به.

يشكل مجمع الوثائق بيانا من العقد المرتبطة فيما بينها، لكن هذا البيان ليس وحيدا دائما، فقد يتكون من اجتماع عدد من البيانات الفرعية. قد تسبب هذه البنية مشكلة من ناحية إمكانية الوصول إلى جميع البيانات انطلاقا من عقدة معينة. لذلك يجب أن تتاح الإمكانية للمستخدم كي يصل إلى أي عقدة من عقد مجمع الوثائق من خلال التصفح. إن بقاء بيان جزئي دون أي روابط تربطه بغيره هو سيئة ينبغي تجاوزها، واكتشاف هذا الأمر قد لا يكون بالأمر السهل، مما يستدعي وجود أداة خاصة ضمن نظام الوسائط الفائقة يقوم بهذا العمل.

2-4-4-2 الضياع

تعاني نظم الوسائط الفائقة من مشكلة أساسية بسبب بنيتها التشعبية هي مشكلة الضياع (Disorientation). ففي كثير من الأحيان يضيع المستخدم أثناء تصفحه ضمن قاعدة الوثائق. وهذا ناتج عن اتباعه لعدد كبير من الروابط، فيجد نفسه في مكان لا يعرف موقعه ضمن قاعدة الوثائق. وفي وفي أحيان أخرى، قد يعاني من مشكلة الإشباع الفكري (Cognitive Saturation)، فكثرة تجواله وملاحقته للمعلومات التي تصادفه، قد تؤدي إلى تشتت ذهنه (Distraction)، ونسيانه للمعلومات التي يبحث عنها.

2-4-4-3 النمذجة

تفتقر نظم الوسائط الفائقة إلى وجود نماذج (Models) عامة لها، وهذا يسبب مشاكل كثيرة من ناحية تصميم هذه الأنظمة وتنفيذها. يجب على أي نموذج عام للوثائق الفائقة أن يحل مشاكل من نمط

مشكلة تعريف المعلومات المطابقة لحاجة المستخدم، فعند إنشاء روابط بين العقد، يجب أن تعكس هذه الروابط العلاقات الموجودة بين العقد.

من ناحية أخرى، يجب على النموذج أن يسمح بتصميم فعال للأنظمة التي تدعم العمل الجماعي المشترك، وأن يسمح للمستخدم بأن يكون بنية الوثائق التي يتصفح ضمنها بالشكل الذي يناسبه، كما يجب أن يدعم إمكانية وجود إصدارات مختلفة من الوثائق، دون أن يؤثر ذلك على إمكانية التصفح.

أما من ناحية الاستخدام، فيجب على النموذج أن يسمح للمستخدمين بالاستخدام المشترك والمتزامن. كما يجب أن يؤمن أدوات من أجل إدارة الوثائق وتأمين سلامة القاعدة والارتباطات، مثل التحقق من إمكانية الوصول إلى جميع العقد ضمن قاعدة الوثائق انطلاقاً من أي عقدة.

2-4-5 عرض لبعض نظم الوسائط الفائقة

تتنوع أنظمة الوسائط الفائقة من أنظمة بسيطة لعرض المعلومات على الشاشة مع تزويدها بروابط للتصفح، إلى أنظمة معقدة تشكل وسطاً متكاملاً من أجل العمل الجماعي المشترك. فيما يلي سنعرض بعض نظم الوسائط الفائقة بهدف تعريف المستخدم ببعض أنواعها وكيفية تطبيق التقنيات ضمنها.

إن النظام الأول الذي سنعرضه هو HyperCard باعتبار أن ظهوره كان نقطة علام في تاريخ نظم الوسائط الفائقة، ثم سنعرض نموذج Dexter كمثال على محاولات نمذجة نظم الوسائط الفائقة لما له من أهمية من ناحية الأفكار التي قدمها و من ناحية استخدامه كنموذج مرجعي لتقييم أنظمة الوسائط الفائقة. ثم سنتكلم عن محرر النصوص Grif الذي يعتمد على مفاهيم الوسائط الفائقة، ونظام MutiCard المستوحى من نظام HyperCard. ونترك التحدث عن نظام الوسائط الفائقة العالمي WWW للجزء الثاني من هذا الكتاب.

2-4-5-1 نظام HyperCard

غير ظهور النظام HyperCard مجال استخدام نظم الوسائط الفائقة. فقبل عام 1987، وهو عام ظهور هذا النظام على يد بيل أتكنسون (Bill Atkinson)، كان استخدام نظم الوثائق الفائقة مقتصرا على الباحثين في عدد من المخابر العالمية (مثل Xerox Parc., MIT). وبعد عدة شهور من تسويقه، تمكن عدة آلاف من المستخدمين من استخدامه.

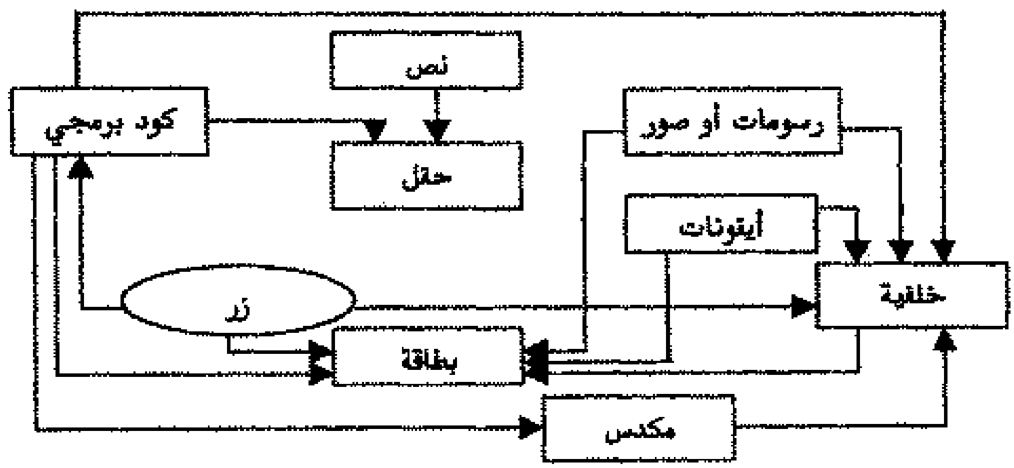
يمكن تعريف HyperCard على أنه نظام توليد وإدارة لنظم الوسائط الفائقة من أجل تخزين المعلومات، والبحث عنها بالتصفح وباستخدام طلبات بسيطة (مطابقة النماذج). يتألف HyperCard من خمسة عناصر هي: المكدس (Stack) والخلفية (Background) والبطاقة (Card) والحقل (Field) والزر (Button).

يستخدم HyperCard مفاهيم البرمجة غرضية التوجه (Object Oriented Programming) في تصميمه: الطرائق (Methods)، والرسائل (Messages) بين الأغراض، والصفوف (Classes) والوراثة (Inheritance). طبقت هذه المفاهيم على عدة مستويات ومن خلال لغة البرمجة الخاصة به وهي HyperTalk.

المكدس هو مجموعة من المعلومات المنظمة. يخزن كل مكدس في ملف مستقل، ويمكن للمكدس الواحد أن يحوي مجموعة من الخلفيات. تحتوي الخلفية المعلومات الموزعة بين مجموعة من البطاقات. تشكل البطاقة الوحدة الصغرى التي يمكن لها أن تحوي أنماطا مختلفة من المعلومات (وهي تقابل العقد حسب مصطلحات نظم الوسائط الفائقة)، كما يمكن لها أن تحوي حقولا وأزرارا. الحقول هي كائنات لا تحتوي إلا معلومات نصية. الأزرار هي أدوات التصفح الأكثر أهمية ضمن المكدس. يبين الشكل (2-4) البنية التصميمية لنظام HyperCard.

يمكن للكائنات (وهي في HyperCard مرتبة هرميا بالشكل التالي: المكدس، الخلفية، البطاقة، الحقل، الزر) أن تتصل باستخدام برامج بلغة HyperTalk ضمن نظام ماكنتوش (نظام التشغيل الذي يعمل عليه HyperCard).

إن الجزء البرمجي موجود في كل غرض، وهو يحدد له كيفية تصرفه تجاه الرسائل التي ترده، يمكن تفعيل كود برمجي معين من خلال مجموعة من الأحداث التي تولد الرسائل مثل الضغط على الفأرة أو كتابة حرف ما، ... وإذا كان الغرض الحالي لا يعالج نوع الرسالة الواردة، فإن هذه الرسالة يجري تحويلها إلى الغرض التالي، وذلك حسب البنية الهرمية للأغراض، ليقوم بمعالجتها.



الشكل (2-4): البنية التصميمية لنظام HyperCard

2-4-5-2 نموذج Dexter

في عام 1988، قامت مجموعة من المصممين بقيادة الباحثين جون ليغيت (John Legett) وجان والكر (Jan Walker) بتشكيل فريق عمل هدفه الوصول إلى تعريف مصطلحات الوسائط الفائقة، وإيجاد تفسير دلالي لمفاهيمها. وقد نتج عن عمل هذا الفريق نموذج يمكن اعتباره نموذجا مرجعيا لتقييم نماذج الوسائط الفائقة. إذن فالغاية من عرض هذا النموذج هي الحصول على المفاهيم المجردة المختلفة الموجودة ضمن نظم الوسائط الفائقة (أو التي يمكن أن توجد فيما بعد).

يتعامل هذا النموذج مع أنظمة الوسائط الفائقة على أنها مؤلفة من ثلاث طبقات (انظر الشكل 2-5):

أ- طبقة التنفيذ (Run-Time Layer): وهي الطبقة المسؤولة عن التخاطب مع المستخدم، من ناحية إدخال الوثائق وعرضها، وتلقي الأوامر من المستخدم. من الصعب جدا أن نجد

نموذجا عاما لجميع بروتوكولات التخاطب التي يمكن استخدامها في هذه الطبقة، مثل طرق الولوج والعرض والتخاطب ومعالجة شبكات الوثائق، لذا لا يعالج نموذج ديكستر سوى الحالات البسيطة من التخاطب ضمن أنظمة الوسائط الفائقة.

ب- طبقة التخزين (Storage Layer): وهي توصف قاعدة البيانات المستخدمة من أجل تخزين الوسائط الفائقة: تخزين البنية الهرمية للعقد، والروابط بين هذه العقد، وبنية شبكة الوسائط الفائقة الناتجة. إن العقد المستخدمة هي عقد ذات توصيف عمومي، فلا تختلف عقد الرسومات عن عقد النصوص.

ج- طبقة محتوى الوثائق (Within-Document Layer): وهي طبقة تحتوي معلومات عن بنية العقد الداخلية وعن محتواها. لم توصف هذه الطبقة في النموذج (عمدا)، لأنه من غير الممكن أن نجد نموذجا يستطيع أن يغطي جميع الحالات (المحتوى/البنية) التي يمكن وجودها في عقدة (نصية أو رسومية أو صوتية أو غيرها). يتعامل النموذج مع معلومات البنية كما لو كانت غير تابعة لنموذج نظام الوسائط الفائقة، لذا فهي توصف خارج النموذج.

يعرف النموذج بروتوكولي اتصال بين الطبقات الثلاث وهما:

- بروتوكول توصيف الروابط (Anchoring): وهو يؤمن الواجهة بين طبقة التخزين وطبقة محتوى الوثائق. يسمح هذا البروتوكول بعنونة مناطق معينة أو مفردات موجودة ضمن طبقة محتوى الوثائق بحيث يصبح بالإمكان الانتقال إليها من طبقة التخزين، مع مراعاة بقاء الفصل بين الطبقتين.

- بروتوكول توصيف العرض: وهو يؤمن الواجهة بين طبقة التخزين وطبقة التنفيذ. يوصف هذا البروتوكول طريقة تخزين العقد والروابط المتعلقة بها بحيث يمكن إظهارها للمستخدم على الشاشة، مع مراعاة بقاء الفصل بين الطبقتين.

وبغية توحيد المصطلحات، سميت الوحدة التي يجري عرضها في طبقة التنفيذ "بمركبة عرض" (Component). يمكن لمركبة العرض أن تكون عقدة أو رابطا أو مركبة عرض مركبة. يعطى لكل مركبة عرض رقم وحيد ومميز على مستوى النظام، كما يزود النموذج بأداتين من أجل الحصول على الرابط انطلاقا من رقمه وبالعكس. يؤمن النموذج أداة لعنونة مناطق معينة ضمن مركبات العرض. تنشئ هذه الأداة مركبة خاصة هي مركبة العنونة وتحتوي قيمتين: رقم مميز لها (لتمييزها عن بقية مركبات العنونة المرتبطة بمركبة العرض)، وقيمة تحدد المكان أو الموقع أو الكلمة التي نريد عنونتها ضمن مركبة العرض. ترتبط إذن بمركبة العرض المعلومات التالية: مجموعة من مركبات العنونة، وموصفات تشرح طريقة العرض، وثنائيات (موصفة/قيمة) تفيد في ربط أي موصفة بمركبة عرض (مثل ربط كلمات مفتاحية بمركبة عرض).

طبقة التنفيذ عرض للوثائق، وتخاطب مع المستخدم.
بروتوكول توصيف العرض
طبقة التخزين قاعدة معطيات تحوي شبكة العقد والروابط
بروتوكول توصيف الروابط
طبقة محتوى الوثائق المحتوى/البنية الخاصة بالعقد

الشكل (2-5): طبقات نموذج ديكستر

تعرف طبقة التخزين، إضافة إلى نموذج البيانات، مجموعة من الوظائف مثل:

- تعديل الوسائط الفائقة ؛
- إضافة أو حذف مركبة عرض ؛

- إنشاء رابط ؛
- إيجاد مركبة عرض أو تعديلها ؛
- تعديل أو إيجاد قيمة إحدى الموصفات ؛
- إيجاد موصفات مركبة عرض ؛
- إيجاد الروابط المرتبطة بمركبة عنونة أو بمركبة عرض.

تعتبر وظيفة نسخ مركبة عرض هي الأهم في طبقة التنفيذ، حيث يقوم النظام بإحضار نسخة عن مركبة العرض، ويعرضها للمستخدم ويسمح له بتصفحها وتعديلها، دون أن يؤثر ذلك على المركبة في طبقة التخزين، لأن النسخة المعدلة تحفظ في طبقة التنفيذ.

من ناحية أخرى، تتضمن طبقة التنفيذ مفهوم "الجلسة" (Session)، الذي يسمح بالمحافظة على التوافق بين مركبة العرض ونسخها. يمكن لطبقة التنفيذ أيضا أن تدير عمليات مثل فتح أو إغلاق جلسة عمل، وفتح أو عرض عنصر عرض، واتباع ارتباط معين، وحذف أو تعديل نسخة مركبة عرض، وتخزين نسخة عن مركبة عرض.

يعتبر نظام Dexter أقدر من معظم أنظمة الوسائط الفائقة الموجودة حاليا. فهذا النظام يعرف مفاهيم للمستقبل مثل روابط من الدرجة n، وروابط بين الروابط، ومركبة مركبة، وغيرها...

إن التوصيف غير الصوري لنموذج Dexter والذي يحاكي الدقة الناتجة عن التوصيف الصوري، والقدرة الكبيرة على التوليد، تجعل منه أداة غنية دلاليا من أجل وصف وتقييم أي نظام وسائط فائقة.

تكمن نقطة القوة في نموذج ديكستر في الفصل المنطقي بين محتوى الوثائق والبنية الهيكلية الناتجة عن الروابط. إن هذا الفصل ممكن بفضل آلية العنونة التي تسمح بوجود عدة بنى هيكلية (شبكات تصفح) على نفس مجموعة الوثائق (العقد).

3-5-4-2 محرر النصوص Grif

طور Grif في مخابر BULL في غرنوبل / فرنسا، وهو محرر نصوص لكتابة الوثائق المهيكلة، حيث تمثل الوثيقة ضمن النظام حسب بنيتها المنطقية. إن البنية المنطقية المقصودة هنا هي البنية الهرمية للوثيقة (فصل، قسم، مقطع،...) والروابط بين الوثائق (كما في الوثائق الفائقة).

يجري تمثيل البنية المنطقية للوثيقة بتحديد عناصر مثل العنوان والكاتب والخلاصة والأقسام والمقاطع والملاحظات والتعليقات والهوامش ... ويمكن للمستخدم أن يحدد الأنماط المختلفة لبنية الوثيقة. يستخدم هذا النظام لغة التأشير (Standard Generalized Markup Language) SGML من أجل تعريف نمط الوثيقة التي تراد كتابتها. يستطيع المستخدم باستخدام SGML أن يعرف نمط الوثيقة، فيحدد جميع أنماط العناصر وجميع العلاقات البنيوية الممكن وجودها بين هذه العناصر. فمثلا، يمكن أن يعرف تقريرا بأنه يجب أن يحوي عنوانا، و اسم الكاتب، وعددا متغيرا من الأقسام ، وكل قسم له عنوان، ومقطعان على الأقل، ....

يجري تعريف البنية المنطقية من خلال مجموعة من الموصفات على النحو الذي تستخدم فيه موصفات من أجل تحديد المضمون الدلالي لعنصر ما باستخدام SGML. ويجري التصريح عن هذه الموصفات أثناء تعريف نمط الوثيقة.

يحتوي نظام Grif على تطبيق لبعض مفاهيم الوثائق الفائقة، فالنظام يسمح بتجزئة الوثيقة إلى أجزاء، ثم الربط بين هذه الأجزاء بروابط، كما يسمح بالربط بين أجزاء من وثيقة وأجزاء من وثيقة أخرى. يمكن للرابطة أن تكون موصفة أو عنصرا، وفي كلتا الحالتين، يمكن لها أن تكون داخلية (مصدر الرابطة ووجهتها تقعان ضمن نفس الوثيقة) أو خارجية (مصدر الرابطة ووجهتها تقعان في وثيقتين مختلفتين). كما يجري تجميع الروابط حسب أنواعها في مجموعتين: روابط مرجعية وروابط احتواء. تنشئ الرابطة المرجعية علاقة بين عنصرين دون القيام بأي تغيير على العناصر، أما رابطة الاحتواء فتسمح باستبدال عنصر الوجهة بعنصر المصدر. تعرف الروابط المرجعية أثناء تعريف نمط الوثيقة، بينما تعرف روابط الاحتواء من قبل المستخدم حصرا.

إن الفائدة الأساسية من نظام Grif تكمن في أنه أثبت عدم وجود فرق أساسي بين الوثائق الفائقة والوثائق المهيكلة، إذا اعتبرنا الروابط بين العقد على أنها روابط بنيوية.

2-4-5-4 نظام MultiCard

طور نظام MultiCard ضمن مشروع أكبر هو ESPRIT MULTIWORKS، وسوق من قبل شركة EuroClid الفرنسية. ويمكن تعريفه على أنه نظام وسائط فائقة يؤمن مجموعة من الأدوات للتعامل مع النصوص الفائقة، ولغة برمجة (MultiTalk)، إضافة إلى محرر نصوص فعال.

إن الهدف الأساسي من هذا المشروع هو تأمين وسط عمل من أجل تنظيم معلومات تجري معالجتها من قبل برامج خارج النظام. أما الحل المقترح فكان بنية مفتوحة مكونة من مجموعة من الوحدات كما يظهر ذلك الشكل (2-6).

تشكل وحدة صفوف الوسائط الفائقة الأساسية (Hypermedia Basic Class <HBC>) طبقة التخزين في النظام (مقارنة مع نموذج ديكستر). تؤمن هذه الوحدة عملية إدارة شبكة العقد والارتباطات في النظام، كما تؤمن واجهة برمجة للتطبيقات (API) من أجل تطوير برامج نظم فائقة.

يسمح البروتوكول M2000 بالتكامل مع تطبيقات خارجية (محررات نصوص)، من أجل معالجة مضمون الوثائق (المعارف الموجودة ضمن الوثائق). يجب على تلك التطبيقات أن تكون قادرة على استخدام هذا البرتوكول أيضا. يسمح هذا البروتوكول لهذه البرامج بتبادل الرسائل فيما بينها، ومع النظام، مثل رسائل طلبات البحث (الاستعلامات)، ورسائل الأجوبة، ورسائل الأحداث ورسائل الخطأ.

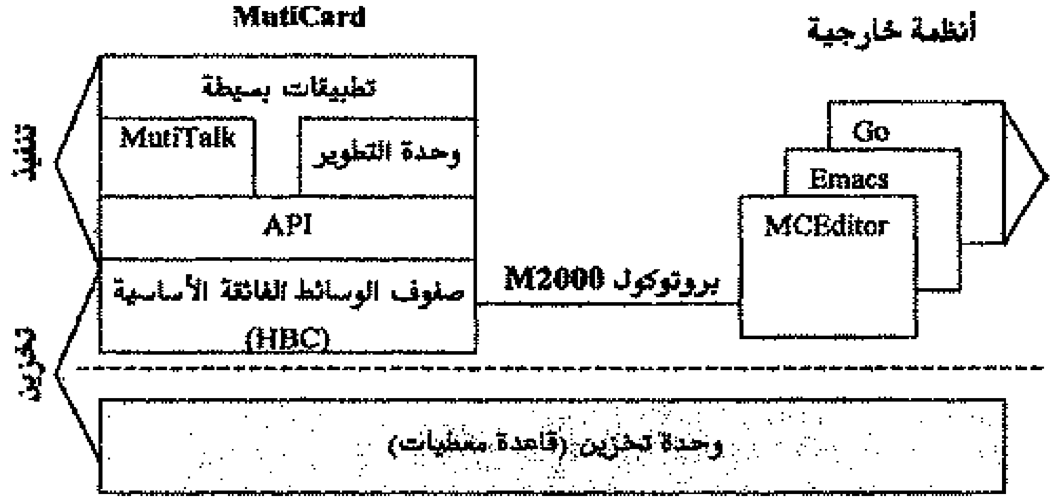
تسمح لغة البرمجة MultiTalk بكتابة برامج بسيطة مرتبطة بالعقد، أو بالروابط، ويمكن أن تفعل اعتمادا على أحداث (Events) مثل إظهار وثيقة أو اتباع رابطة، ... تحتوي هذه اللغة على 150 تعليمة تسمح بمعالجة محتوى العقد، وتزامن المحررين أثناء عملهم.

تسمح وحدة التخزين بتخزين المعلومات بشكل موزع على عدة أجهزة. ويجري حفظ المعلومات وإدارتها من قبل نظام قواعد معطيات، وتمكن بنية مخدم/زبون التطبيقات من أن تصل إلى المعلومات المخزنة على عدة أجهزة ضمن الشبكة. إن نظام إدارة هذه المعلومات ليس نظام قواعد معطيات كاملا، إذ تنقصه مجموعة من الوظائف مثل معالجة حالات الأعطال، ولغة الاستعلامات.

إن عملية تخزين الوثائق هي من مسؤولية محرر النصوص المستخدم، فعندما ينشئ المستخدم عقدة جديدة، يجب عليه أن يحدد أحد المحررات من أجل عملية إدارة وتنظيم محتوى هذه العقدة، ومن ثم عملية تخزينها ؛ إذن عملية التخزين وطريقة العرض تتعلقان بالمحرر. إن فكرة الفصل بين محتوى العقد وطريقة عرضها (التي يقوم بها برنامج المحرر) تعد من الأفكار الهامة التي يقدمها هذا النظام، لأنها تبقي المعلومات محدثة دوما، إذ يكفي تغيير طريقة تفسير مخطط عرض العقدة من أجل تحديث طريقة العرض.

يعتبر MutiCard الوثائق الفائقة كبيان (Graph) عقده هي إما عقد بسيطة تحوي معلومات متعددة الوسائط، وإما عقد مركبة يكون محتواها مجموعة من العقد. ولا يوجد أي قيد على الروابط بين عقد النظام، فيمكن للرابطة أن توجد بين عقدتين من نفس النوع أو عقدتين من نوعين مختلفين (عقدة بسيطة وعقدة مركبة). تمتلك كل عقدة أو رابطة ضمن النظام مجموعة من الموصفات، وتتألف كل موصفة من ثنائية (اسم/قيمة) (راجع نموذج ديكستر). يمكن تعديل هذه الموصفات أو حذفها من خلال لغة البرمجة MutiTalk.

إن أشكال التكامل بين النظام والمحررات الموجودة متنوعة جدا، ففي أبسط الحالات، يمكن أن يستخدم المحرر ليقوم بعمليات الإظهار وحفظ العقد، وفي مستوى تكامل أعلى، يمكن للنظام أن يسمح لمحرر النصوص بأن ينشئ روابط كما يلي: يحدد المستخدم منطقة داخل العقدة، ويطلب من المحرر أن ينشئ عنوانا لهذه المنطقة. يقوم المحرر بالاتصال بالنظام (رسالة) الذي سيقوم بعملية إنشاء العنوان ضمن العقدة المحددة، ومن ثم يمكن للمستخدم استخدام هذا العنوان كمصدر أو وجهة للرابطة التي يريد إنشاءها.



الشكل (2-6): البنية التصميمية لنظام MutiCard

2-4-6 اعتبارات بناء نظام وثائق فائقة والتقنيات المتبعة

عند بناء نظام وسائط فائقة يجب الانتباه إلى المجال الذي سيطبق النظام فيه، وعلى ضوء هذه المعلومة يجري التركيز على أجزاء النظام. فمثلا في حالة الأنظمة المستخدمة لأغراض تعليمية، يجب إفساح المجال أمام المستخدم لكي يحدد طريقة التصفح التي يريد اتباعها ليستعلم عن الموضوع المطروح، وبالتالي يجب التركيز على الروابط بين العقد أكثر من التركيز على البنية الهيكلية (المنطقية) للوثائق والمقترحة من قبل مصمم النظام، كما يجب الانتباه إلى أن المستخدم قد يحتاج إلى إضافة تعليقات خاصة به. أما في الأنظمة المستخدمة في عمليات إدارة المعلومات واتخاذ القرارات، فيجب التركيز على البنية المهيكلة للوثائق أكثر من التركيز على الروابط، لأن المطلوب هو الوصول إلى معلومات معينة وفق طريقة معينة، أكثر منه التعلم، كما أن إمكانية إضافة الملاحظات متوفرة. وفي أنظمة العرض، مثل أنظمة عرض الوثائق التقنية أو دليل الاستخدام، يجري التركيز على البنية المهيكلة، إضافة إلى الفصل بين نظام الكتابة (النظام الذي استخدم من أجل إعداد هذه الوثائق) ونظام القراءة والتصفح (النظام المستخدم من أجل تصفح وقراءة هذه الوثائق)، إن نظام المساعدة (Help)

المرفق مع معظم البرامج والذي يشرح طريقة استخدام هذه البرامج هو مثال جيد عن هذه الحالة الأخيرة، فهو يسمح بتصفح واستعراض الوثائق دون السماح للمستخدم بأن يعدل عليها. وفي الحالات التي نريد استخدام النظام كوسط متكامل للعمل المشترك، كما في الحالات التي نصادفها في هندسة البرمجيات أو معالجة المعلومات حيث نحتاج إلى وسط يسمح لعدد من المستخدمين بالتعاون فيما بينهم لإنجاز مشروع مشترك، يجري التركيز على العقد والروابط معا، مع إمكانية إضافة الملاحظات والتعليقات.

يجب الانتباه إلى أمر آخر وهو عدد مستخدمي النظام، فأنظمة الوسائط الفائقة قد تكون موجهة إلى مستخدم واحد أو مجموعة كبيرة من المستخدمين كما في حالة مجموعة من المبرمجين الذين يعملون بالتعاون فيما بينهم، كما يمكن لنظام الوسائط الفائقة أن يعالج مجموعة من الوثائق بحجوم قد تكون ضخمة جدا. إن هذه الأمور يجب أن تؤخذ بعين الاعتبار عند الحديث عن التخزين وتصميم النظام، كما تؤثر على طريقة استخدام النظام والواجهات. ففي الأنظمة التي سيستخدمها عدد كبير من المستخدمين، يجب السماح لهؤلاء جميعا بأن يصلوا إلى أي معلومة، دون الاضطرار إلى الانتظار كثيرا، كما يجب ألا يعيق بعضهم بعضا أثناء عمليات التصفح والبحث.

2-4-7 البحث عن المعلومات

تؤدي نظم الوسائط الفائقة إلى طريقة جديدة في البحث عن المعلومات من حيث المنهج، فنظم قواعد المعطيات ونظم استرجاع المعلومات تدعم اتجاه البحث عن المعلومات باستخدام الطلبات، حيث يقوم المستخدم بصياغة المعلومات التي يريدها على شكل طلب (استعلام)، ثم يطلب من النظام البحث عنه، أما النظم الخبيرة فهي تدعم عمليات البحث بالاستدلال، فهي تستخدم قاعدة حقائق في عمليات البحث عن المعلومات التي يريدها المستخدم.

تدعم نظم الوسائط الفائقة البحث بالتصفح الحر، أي التصفح ضمن الوثائق وتجميع المعلومات. هذا التصفح والتوجيه محكوم بأمور متعددة منها ما يعود إلى طبيعة المستخدم الذي يقوم بالتصفح وطريقة

ربطه للأمور، ومنها ما يعود إلى إمكانيات هذه النظم من حيث طريقة توزيعها للمعلومات في عقد وطريقة تعريف الروابط بين هذه العقد. وتجدر الإشارة هنا إلى أن معظم نظم الوسائط الفائقة تتيح نوعا من البحث البسيط يعتمد على مطابقة النماذج (Pattern Matching) من نمط البحث عن كلمة أو عن مجموعة كلمات في نصوص العقد.

5-2 البحث عن المعلومات في النظم الهجينة

نسمي نظم المعلومات التي تتيح جميع (أو معظم) إمكانات البحث التي تمتع بها النظم الثلاثة السابقة النظم الهجينة، هذا وتعتبر معظم نظم البحث الحديثة من هذا النمط. ومن الطبيعي أن تصميم هذه النظم يخضع للتصميم المعروض في الشكل (2-1)، أي أن تخزين المعطيات يجري في نظام إدارة قواعد البيانات (مما يسمح بالبحث عن موصفات البيانات)، وتبنى فوقها مركبتان: واحدة لإتاحة إمكانيات نظم استرجاع المعلومات (أي البحث عن محتوى المعلومات)، وأخرى لإتاحة إمكانيات نظم الوسائط الفائقة (أي التصفح الحر والتنقل بين المعلومات المختلفة). مما لا شك فيه أن مكاملة النظم الثلاثة في نظام واحد تجري غالبا بأشكال مبسطة، إذ أنه من الصعب جدا استيعاب جميع إمكانيات هذه النظم في نظام عملي واحد.

تعتبر محركات البحث (Search Engines) في الإنترنت أهم الأمثلة على هذه النظم، وهي أيضا هدفنا من هذا الكتاب، لذلك فقد خصصنا لها القسم الأكبر من الجزء الثاني منه.

الفصل الثالث

تقييم ومستقبل البحث عن المعلومات

ركزّنا الاهتمام في الفصل السابق على نظم استرجاع المعلومات ونظم الوسائط الفائقة، وفي هذا الفصل سنقيّمَ هذين النوعين من نظم البحث عن المعلومات، ونضع بعض التصورات المستقبلية لكلٍ منهما على حدة، ثم نقارن بينهما وفق بعض المعايير، لنخلص أخيراً إلى وضع تصوّر لمستقبل نظم البحث عن المعلومات.

3-1 تقييم نظم استرجاع المعلومات

يمكن تلخيص مميزات نظم استرجاع المعلومات في النقاط التالية:

- استخدام التمثيل المعرفي من أجل البحث عن المحتوى، وإن كانت درجة تعقيد هذا التمثيل تختلف من نموذج لآخر.
- غالباً ما تكون الواجهات التخاطبية لهذه النظم ضعيفةً، ولا تسمح إلاّ بكتابة الطلبات كطريقة وحيدة لإيجاد المعلومات. ويُعتبر هذا ضعفاً كبيراً من وجهة نظر التخاطب بين الإنسان والآلة: عندما يجيب النظام عن طلبٍ لمستخدم نهائي، فإن هذا المستخدم لا يرى إلاّ هذه الإجابة، وليس لديه إمكان آخر سوى كتابة طلب جديد، فكيف يمكن إذن تحسين

هذه الإجابة ؟ هذا يعتمد على خبرة المستخدم (بلغة الفهرسة المتبعة في فهرسة المعلومات المخزنة)، وغالباً ما يؤدي ذلك إلى إجراء تخاطبي طويل، وفي نهايته لن يعرف هذا المستخدم أبداً إن كان قد استطاع إيجاد جميع المعلومات الموائمة لطلبه، أم لا.

- لا توجد مشكلة ضياع في نظم استرجاع المعلومات، لأن المستخدم النهائي يُطالب دائماً بتحديد حاجته من المعلومات (أي كتابة طلب صريح).

- تتعامل نظم استرجاع المعلومات عموماً مع معلومات عنصرية (غير مجزأة). وهي تتجاهل بنية المعلومات، فمثلاً تتعامل مع الوثيقة ككتلة واحدة. ولا يوجد هنا استرجاعٌ الفقرات مثلاً، مع أن ذلك يجب أن يؤخذ بعين الاعتبار: وهنا قد يتصور المرء أن استراتيجية البحث ستتغير تغيّراً دراماتيكياً وغير متوافق البتة مع التعريف الحالي لنموذج استرجاع المعلومات. لنأخذ مثالاً بسيطاً: نفترض أنّه لدينا الطلب Q ولدينا فهرسة منفصلة للمقاطع والفقرات في الوثائق. يستطيع النظام إذن إيجاد المقطع S والفقرة P كإجابات موائمة للطلب Q. فماذا يحدث إذن إذا كانت P جزءاً من S (P فقرة من المقطع S): الإجابة ستكون مكرّرةً وقد يعتبرها المستخدم مضللةً له.

- صُمم الكثير من نماذج استرجاع المعلومات لتتعامل مع منظومات معقدّة لتمثيل المعارف (كمنظومة البيانات المعرفية المستخدمة في نظام ELEN وفي نظام HyperRime).

- حساب المطابقة يعتمد على نماذج رياضية متينة وإن كانت تنقصها بعض التبريرات. مثلاً لماذا القياس Tf/idf في النموذج الشعاعي أو تجيب الزاوية بين الطلب والوثيقة ؟.

- بناء نظام استرجاع معلومات يتطلب الكثير من النمذجة الرياضية والعمل الخوارزمي والبرمجي.

2-3 تقييم نظم الوسائط الفائقة

يمكن تلخيص مميزات نظم الوسائط الفائقة في النقاط التالية:

- نظم الوسائط الفائقة هي نظم أليفة الاستخدام (Friendly)، ومزودة بواجهات تخاطبية "جميلة" ولا يتطلب استخدامها عموماً أي خبرة خاصة بها.
- يبرهن مفهوم التصفح على إمكان استخدام هذه النظم في تنظيم المعلومات واسترجاعها.
- للتصفح في الوسائط الفائقة حدوده أيضاً، فالمستخدم يهدر الكثير من الوقت في لبحث بطريقة التجربة والخطأ (Trial-and-Error)، وغالباً ما يتعرض المستخدمون للمشكلة المشهورة "بالضياع": يضيع المستخدم النهائي بعد انقضاء مدة من البحث في شبكة المعلومات، ويصبح بحاجة إلى أن يعرف أين هو، وإلى أين سيذهب، وكيف يذهب.
- تتعامل نظم الوسائط الفائقة مع معلومات مهيكلة إلى حد بعيد. فعلى سبيل المثال، التجول في وثائق نصيّة مهيكلة طبيعي وبسيط: التجول في البنية المنطقية للوثيقة من الفصول إلى المقاطع إلى الفقرات، فضلاً على إمكان استخدام الروابط التشعبية كالهوامش والروابط مع الوثائق الأخرى.
- تستخدم نظم الوسائط الفائقة غالباً طرقاً بسيطةً لتمثيل المعارف (كالكلمات المفتاحية)، لأنّه من المتعذر بناء روابط اعتماداً على معارف متقدمة، ومن ثم لمساعدة المستخدم النهائي على الوصول إلى المعلومات التي يبحث عنها بواسطة شبكة الوسائط الفائقة.
- بناء نظام وسائط فائقة يتطلب الكثير من النمذجة لواجهات التخاطب مع المستخدم، ولطرق توزيع المعلومات في عقد، ولكيفية بناء الروابط بين هذه العقد.

3-3 بعض التصورات المستقبلية لنظم استرجاع المعلومات

- مشاكل التمثيل المعرفي لمحتوى الوثائق أصبحت شبه محلولة بالنسبة للنصوص، أما بالنسبة للوسائط الأخرى كالصور والفيديو والأصوات فإنها باب واسع للبحث العلمي حالياً.
- تطوير نماذج الاسترجاع باتجاه النماذج النظرية والتمثيل غير المؤكد للمعارف ونحو استيعاب البنى المنطقية للوثائق وخصوصيات التجول على نحو أفضل.
- تطوير الواجهات التخاطبية باتجاه تسهيل عملية بناء الطلبات.

4-3 بعض التصورات المستقبلية لنظم الوسائط الفائقة

- تطوير آليات البحث في نظم الوسائط الفائقة باتجاه استيعاب وظائف ونماذج نظم استرجاع المعلومات مع الحفاظ على عامل السرعة.
- الحساب الآلي للروابط بين مختلف أنواع عقد المعلومات.
- التمثيل الدلالي لمحتوى المعلومات واستخدام قواعد المعرفة في البحث.
- نمذجة المستخدم والواجهات التخاطبية.
- إدارة ونمذجة آليات إضافة/حذف/تعديل المعلومات في هذه النظم.
- استخدام تقنيات تابع المطابقة في نظم استرجاع المعلومات، من أجل تثقيل الروابط بين الوثائق في نظم الوسائط الفائقة، وبالتالي التحسين من نوعية التصفح.

5-3 مقارنة نظم استرجاع المعلومات مع نظم الوسائط الفائقة

نقوم فيما يلي بإجراء مقارنة سريعة بين هذين النوعين المختلفين من النظم، ونبيّن كيف يمكن مكاملتهما من أجل الوصول إلى بحث عن المعلومات أكثر فاعلية ممّا يمكن الحصول عليه باستخدام كلٍّ منهما على حدة. نجري المقارنة وفق المعايير العشرة التالية:

* من ناحية النماذج وجدنا أن نماذج نظم استرجاع المعلومات أمتن رياضياً.
* من ناحية الواجهات لا شك أن واجهات نظم الوسائط الفائقة أفضل بكثير من واجهات نظم استرجاع المعلومات، وخاصيّة التصفح قد طوّرت التفاعل بين الإنسان والآلة كثيراً.
* من ناحية تمثيل المعارف فهي موجودة على نحوٍ مدروس أكثر في نظم استرجاع المعلومات، وسبب ضعفها في النوع الثاني يعود لاعتبارات تتعلق بسرعة البحث وزمن الاستجابة.
* من ناحية ترتيب النتائج فإن معظم نظم استرجاع المعلومات تتيح هذه الإمكانية، في حين مازالت في النوع الثاني محدودةً جداً.
* من ناحية جودة النتائج فلا شك أنها أفضل في نظم استرجاع المعلومات بسبب تمثيل المعارف ودقة بناء الطلبات.
* من ناحية بناء الطلبات فإنها تعتبر أفضل بقليل في نظم استرجاع المعلومات، وهذا يعود إلى الحرص على دقة الطلبات فيها.
* من ناحية التغذية الراجعة فإنها مفيدة أكثر في نظم استرجاع المعلومات بسبب تعقّد نماذجها.
* من ناحية إمكانيات التصفح فمن الطبيعي أن نظم الوسائط الفائقة أفضل بكثير لأنها أصلاً مبنية على التجول.

- من ناحية التعامل مع البنية فإن ذلك أفضل بقليل في نظم الوسائط الفائقة، لأنها تتعامل مع بنية المعلومات بفاعلية أكبر.
- وأخيراً من ناحية الفلترة التي هي أصلاً عملية مستقاة من قواعد البيانات (البنية التحتية لكلا النوعين) فإنها متقاربة في النوعين.

3-6 مستقبل نظم البحث عن المعلومات

نلاحظ من المقارنة السابقة أنه عندما يكون المعيار ضعيفاً في أحد النوعين فإنه يكون جيداً في النوع الآخر، وهذا ما يجعلنا نعتقد فعلاً أن الحل الأمثل هو مكاملة هذين النوعين من طرق البحث عن المعلومات. وهذا ما حدا بمحركات البحث عن المعلومات في الإنترنت إلى الاستفادة من إمكانيات كلا النظامين، إضافةً إلى بنائهما فوق نظام إدارة قواعد بيانات، مما يؤدي أيضاً إلى الاستفادة من إمكانيات قواعد البيانات في البحث عن موصفات الوثائق وعمليات الفلترة.

من كلّ ذلك نستنتج أن مستقبل نظم البحث عن المعلومات هي نظم شبيهة بمحركات البحث على الإنترنت وبحيث تتطوّر لتشمل فعلياً جميع خواص نظم البحث عن المعلومات المختلفة، وستكون عملية المكاملة هذه في جميع مستويات النظام (أي من وضع النموذج الرياضي وصولاً إلى مستوى الواجهات التخاطبية).

الجزء الثاني

طرق البحث عن المعلومات في الإنترنت

مقدمة الجزء الثاني

نقتطف العبارات التالية للصحفي فانوفار بوش Vannevar Bush من مقالة له بعنوان "كما يمكن أن نفكر" (As we may think) والتي نُشرت في تموز عام 1945 في شهرية الأطلسي عندما كان يعمل مراسلاً لها في جنوب شرق آسيا إبان الحرب العالمية الثانية:

" ... the human mind does not work like that. It works by association.
With one item in its grasp,
to the next that it suggested by the association of thoughts, it snaps instantly
in accordance with some intricate web of trials carried by the cells of the brain.
It has other characteristics, of course ;
trials that are not frequently followed are prone to fade,
items are not fully permanent, memory is transitory.
Yet the speed of action, the intricacy of trials,
the details of mental pictures, is awe-inspiring beyond all in nature ... "

لم يكن بوش يومها مدركاً أنه بمقالته هذه سيؤسس لواحدٍ من أهمّ اختراعات القرن العشرين وهو النص الفائق (Hypertext)، ومن ثم الوسائط الفائقة (Hypermedia) التي ترافقت مع انتشار الوسائط المتعددة (Multimedia).

تخصص هذا الجزء للحديث باستفاضة عن طرق وأدوات البحث عن المعلومات في الإنترنت، والتي تعتمد كلها جزئياً (أو كلياً) على مفاهيم الوسائط الفائقة المشروحة في الفصل الرابع. كما سيشمل

الحديث الخدمات الأخرى التي تقدمها هذه الأدوات بهدف التعرف الجيد على كل إمكاناتها مما يحسن نوعيا آليات استخدامها.

ففي الفصل الرابع نعرف شبكة الإنترنت ومفاهيمها الأساسية بسرعة، ثم نشرح مجموعة من أكثر أدوات البحث عن المعلومات في الإنترنت انتشارا.

وباعتبار أن محركات البحث (Search engines) هي أهم أدوات البحث عن المعلومات في الإنترنت فقد خصصنا لها ما تبقى من الكتاب.

وهكذا نعرض في الفصل الخامس تعريف محركات البحث تعريفا عاما، ثم نحدد مجموعة من المعايير المنتقاة من الأدبيات في هذا المجال، إضافة إلى ما استقيناه من تجربتنا الخاصة، بهدف المقارنة بين المحركات المختلفة مما قد يساعد الباحث عن المعلومات على اتخاذ القرارات الصائبة أثناء عمليات البحث.

في الفصل السادس نستعرض أهم محركات البحث العامة:

Yahoo!, HotBot, AltaVista, InfoSeek, Lycos, Excite, WebCrawler.

حيث نركز الاهتمام في خواص كل محرك وفق المعايير المحددة في الفصل الخامس.

في الفصل السابع نستعرض باختصار نوعين خاصين من محركات البحث هما: محركات البحث البينية (Meta Search Engines) ومحركات البحث العربية.

في الفصل الثامن، نقارن بين محركات البحث المختلفة وفق المعايير المحددة في الفصل الخامس.

ونختتم الكتاب أخيرا في الفصل التاسع بخلاصة عامة للجزء الثاني، وببعض النصائح الشخصية حول استخدام محركات البحث، وبآفاق مستقبلية عامة في مجال أدوات البحث عن المعلومات في الإنترنت.

الفصل الرابع

شبكة الإنترنت وخدمات البحث عن المعلومات

لقد خصصنا الكتاب الأول من هذه السلسلة لدراسة شبكة الإنترنت وبنيتها التحتية والتقانات التي تستخدمها والخدمات التي توفرها. سنركز الاهتمام في هذا الفصل بأهم خدمات الإنترنت المفيدة في مجال البحث عن المعلومات.

4-1 تعريف شبكة الإنترنت

تربط شبكة الإنترنت عدداً هائلاً من الحواسيب بعضها ببعض عبر شبكة واسعة تضم عشرات ملايين الأجهزة موزعةً على أغلب دول العالم. ويتزايد عدد الأجهزة المرتبطة في هذه الشبكة ازدياداً مستمراً، كما يتزايد عدد مستخدميها يوماً بعد يوم.

يمكن للحواسيب المتصلة بعضها مع بعض من خلال شبكة الإنترنت أن تتبادل المعلومات فيما بينها، فيمكن لحاسوب ما أن يدخل ويعاين وينسخ المعلومات المخزنة على حاسوب آخر. ندعو الحاسوب الذي نعمل عليه بالحاسوب المحلي (Local Computer)، في حين ندعو الحاسوب الذي نعاين المعلومات عليه بالحاسوب البعيد (Remote Computer).

إذا تصورنا العدد الضخم من الحواسيب المرتبطة فيما بينها بشبكة الإنترنت وتخيلنا كذلك التنوع الهائل للمؤسسات والدوائر والشركات والأفراد الذين يعرضون معلومات على هذه الشبكة، فيمكن عندئذ أن نتخيل الكم الهائل من المعلومات الذي يمكننا أن نحصل عليه عن طريق الارتباط بشبكة الإنترنت.

4–2 تاريخ الإنترنت

اقترح بول باران Paul Baran في عام 1962 نظاماً من الحواسيب المتصلة بعضها ببعض ويشمل الولايات المتحدة كافة، وذلك باستخدام شبكة لامركزية بحيث تولف بقية العقد اتصالاتها ديناميكيا إذا تعطلت أو انشغلت بعض العقد.

في عام 1969 تأسست أول شبكة لتبديل الطرود (Packet-Switching network) من قبل وكالة مشاريع الأبحاث المتقدمة ARPA التابعة للبنتاغون، وقد سميت آنذاك بـ ARPAnet. ضمت شبكة ARPAnet أربعة مراكز أبحاث: جامعة كاليفورنيا في لوس أنجلوس (UCLA) ومعهد أبحاث ستانفورد (SRI) وجامعة كاليفورنيا في سانتا باربارا (UCSB) وجامعة أوتا (Utah).

توسعت شبكة ARPAnet لتضم في نهاية عام 1972 حوالي 40 عقدة. وفي ذلك العام أيضاً ظهرت مجموعة عمل ترابط الشبكات ((Internet Working Group (INWG)) التي انحصرت مهمتها في تأسيس بروتوكولات للنظم التي بدأت تنمو بسرعة كبيرة.

بفضل كونها لامركزية، كان من السهل على مديري النظم إضافة الحواسيب لشبكة ARPAnet. وكل ما كان يتطلبه الأمر هو خط هاتفي وأجهزة قليلة العدد وبروتوكول NCP مجاني. خلال بضع سنوات أصبح هنالك أكثر من مائة حاسوب ضخم (Mainframe) مرتبط بالشبكة التي عبرت البحار.

سرعان ما أصبحت ARPAnet مكاناً لتبادل المعلومات والأفكار وقد كان التعاون بين العلماء والمدرسين هو مجال الاستخدام الأكثر شيوعاً للنظام. ولم يكن مستغرباً أن أول تطبيق رئيسي طوّر للاستخدام على الشبكة كان البريد الإلكتروني (Electronic Mail).

مع ظهور نظام البريد الإلكتروني الذي طوره Ray Tomlinson في عام 1972 استطاع الباحثون المرتبطون بالشبكة تبادل الأفكار والأبحاث بسهولة لم يكن من الممكن تخيلها سابقاً. بإضافة إمكانية إرسال البريد إلى عدة أشخاص معاً ظهر ما يعرف بلوائح البريد (Mailing Lists). قاد التطور المنطقي لمفهوم لوائح البريد إلى المؤتمرات التخاطبية أو مجموعات الأخبار (Newsgroups).

في عام 1976 أضافت مخابر Bell في شركة ATCT التطبيق UUCP (Unix-to-Unix Copy) إلى نظام التشغيل UNIX V7. قام بعدها مجموعة من المطورين بتطوير أول نسخة من شبكة Usenet، التي سرعان ما أصبحت بعد ذلك نظام المؤتمرات المباشرة الأمثل على الشبكة. ظهر بعد ذلك البروتوكول NNTP (Network News Transfer Protocol) في عام 1986 لتحسين فعالية شبكة Usenet عبر شبكات TCP/TP. أدى هذا الأمر إلى توسيع الشبكة لتلائم عدد المشتركين فيها والذي جاوز حالياً عدة ملايين في العالم.

لقد كان العديد من المؤسسات الحكومية حتى منتصف السبعينات مرتبطاً بشبكة ARPAnet، ولكن كلاً منها كان يعمل على شبكة طورتها شركة مختلفة. مثلاً، بنت شركة DEC نظام الجيش وبنت IBM نظام القوى الجوية، أما نظام البحرية فقد بنته شركة Unisys. لقد كانت كلها شبكات قوية وفعالة ولكنها كانت تتخاطب بلغات مختلفة. لقد كان من الواضح أن جعل الأمور تجري على نحو أفضل وأدق يتطلب مجموعةً من البرتوكولات الشبكية (Protocols) تربط الشبكات المتنافرة بعضها ببعض وتسمح لها بالتخاطب فيما بينها.

ظهر البرتوكول TCP/IP في عام 1982 كحل للمشكلة السابقة، ومع تطور TCP/IP دخلت كلمة الإنترنت (Internet) إلى اللغة (Internet هي اختصار للتعبير Interconnected Networks).

يتمتع البروتوكول TCP/IP بميزات عديدة نذكر منها إمكانية استخدامه لاتصال الحواسيب على شبكة محلية أو لتخاطب حواسيب من شبكات محلية مختلفة، فهو يسمح بربط الشبكات فيما بينها، كذلك يتحقق TCP/IP من البيانات المرسلة بين حاسوبين ويكتشف الأخطاء فيها أو ضياع جزء منها أو تكرار إرسال بعضها.

أدت الميزات الكبيرة للبروتوكول TCP/IP إلى إعلان وزارة الدفاع عن اعتماد طقسم بروتوكولات TCP/IP كبروتوكول قياسي لربط الحواسيب العسكرية. وقد أنشئت نسخ عنه تعمل على أنظمة مختلفة تتضمن الحواسيب الشخصية والحواسيب الضخمة (Main Frames) وغيرها، وأصبح القياس الجديد لتشبيك الحواسيب. وهو يكوّن اليوم مجموعة البرتوكولات التي تقدم البنيسة التحتية لشبكة إنترنت.

لقد كان من أهداف تطوير ARPAnet تأمين إمكانية دخول الباحثين إلى حاسوب بعيد وتنفيذ البرامج عليه. في ذلك الوقت كانت توجد بعض الحواسيب والقليل جدا من الحواسيب الفائقة (لم تكن الحواسيب الفائقة في أعوام السبعينيات ترقى إلى قوة أي حاسوب شخصي اليوم). لقد كانت إمكانية الدخول عن بعد (Remote logon) إضافةً إلى البريد الإلكتروني أولى الإمكانيات التي بنيت على شبكة ARPAnet.

لم يعد هذا السبب قائماً اليوم فلا حاجة لدخول المستخدمين إلى حواسيب بعيدة، إذ أن أغلب المؤسسات الحكومية والجامعات ومراكز الأبحاث تملك حواسيبها الخاصة التي تضاهي قوةً وقدرةً الحواسيب الموجودة في الأماكن الأخرى.

يقدم TCP/IP إمكانية الدخول عن بعد من خلال البروتوكول Telnet الذي يسمح لمستخدم يملك حسابا على نظام Unix بالدخول من حاسوب بعيد وتنفيذ البرامج على حاسوب Unix.

تتطلب إمكانية الدخول عن بعد استخدام برنامج مستقل مثل البرنامج Telnet على أنظمة Unix وعلى أنظمة Windows.

مع تحول أغلب الشبكات إلى استخدام البروتوكول TCP/IP في السنوات التي تلت عام 1982 أصبحت كلمة إنترنت هي التعبير الشائع للإشارة إلى الشبكة العالمية التي تربط حواسيب منتشرة في مراكز الأبحاث وفي الجامعات وفي المواقع العسكرية.

في عام 1983 قسمت ARPAnet إلى جزأين: ARPAnet و MILNET، وسرعان ما تكاملت MILNET مع شبكة بيانات الدفاع (Defense Data Network) التي أنشئت عام 1982. بعد ذلك استحوذت شبكة NSFNET (National Science Foundation NETwork) التي أنشئت عام 1986 على دور ARPAnet لتصبح العمود الفقري الشبكي بمساعدة NASA ووزارة الطاقة، وذلك بغية تأمين سرعة اتصال أساسية تصل حتى 56 KBps لتشبيك جيل جديد من الحواسيب العملاقة. بعد ذلك تزايدت ارتباطات الشبكات مع هذه الشبكة المركزية وخاصة مع المراكز التعليمية.

في عام 1993 أنشيء مركز معلومات شبكة إنترنت (Internet Network Information Center <InterNic>) من قبل هيئة العلوم القومية (National Science Foundation) بغية تقديم المعلومات وخدمات الأدلة وقواعد البيانات والتسجيل في مجتمع إنترنت ؛ ويمكن النظر إلى InterNIC على أنه مركز إدارة إنترنت ولكنه لا يملي سياستها ولا يشغّل حاسوباً مركزياً ضخماً للتحكم بها، إذ تنحصر مهمته في الوظائف التنظيمية وحفظ السجلات مثل إسناد عناوين إنترنت فقط.

في عام 1991 أنشأ بول ليندر Paul Linder ومارك ماكاهيل Mark P. McChill من جامعة مينسوتا خدمة غوفر Gopher. وقد لقي غوفر، بفضل تقنيته في تنظيم الملفات تحت نظام قوائم حدسي، القبول مباشرة على الشبكة.

يتعامل غوفر مع جميع البيانات على أنها قوائم أو وثائق أو فهارس أو اتصال Telnet. من خلال Telnet يستطيع موقع غوفر النفاذ إلى مواقع أخرى مما يجعله تطبيقاً شبكياً حقيقياً قادراً على جلب البيانات إلى المستثمر من عدة مواقع عبر واجهة واحدة.

لقد اعتُبِرَ غوفر المبشرَ بولادة شبكة الوب العالمية (World Wide Web) من خلال مفهومه ووظائفه، وإن كانت تنقصه الارتباطات التشعبية (Hyperlinks) والعناصر الرسومية. لقد بدأت خدمة الوب

تحلُّ محله اليوم، وعلى الرغم من أنه توجد حتى الآن مواقع غوفر عديدة فإننا نعتقد أنها ستختفي كليا في السنوات القريبة القادمة.

4-3 خدمات البحث عن المعلومات في الإنترنت

4-3-1 الشبكة العنكبوتية العالمية WWW

الشبكة العنكبوتية العالمية، أو الوب اختصاراً، هي نظام وسائط فائقة موزع على مستوى العالم، يمكن الوصول إليه عن طريق الإنترنت. طُوِّر الوب في مختبرات CERN الأوروبية بهدف تمكين الباحثين في أماكن مختلفة من العالم من تبادل المعلومات والوثائق العلمية.

يسمح الوب للمستخدم بالبحث عن المعلومات بالتصفح في وثائق النظام تصفحاً شفافاً متقصياً عن أماكن تخزين هذه الوثائق في مختلف مخدّمات الوب على شبكة الإنترنت.

ينسجم الوب عموماً مع المفاهيم التي عرضناها في الجزء الأول، لكنه يطبق بعض هذه المفاهيم بما يتناسب مع طبيعته الموزعة، كما يعدّل أحياناً في بعض المصطلحات.

فعقد المعلومات في نظم الوسائط الفائقة أصبحت تُسمى صفحات الوب (صفحات HTML). وقد رأينا في الفصل الرابع أن عقد المعلومات قد تكون على عدة أنواع (عقد نصوص، أو عقد صور أو...)، إلا أن الوب لا يملك سوى نوع واحد من عقد المعلومات. لكن العقدة (الصفحة)، تستطيع أن تحوي أي نمط من المعلومات والوسائط المتعددة. أما الروابط بين عقد المعلومات، والتي توجد في نظم الوسائط الفائقة، فتقابلها الارتباطات التشعبية بين الصفحات في الوب. كما نلاحظ أن هذه الارتباطات التشعبية وحيدة الاتجاه، أي إذا ارتبطت صفحة A بصفحة B، فهذا لا يعني أبداً أن B ترتبط بـ A (وهذا ما كانت عليه بعض أنواع الروابط في نظم الوسائط الفائقة). من ناحية أخرى، هذه الارتباطات التشعبية هي روابط ساكنة، ولا وجود في الوب لما يسمى بالروابط الديناميكية أو لأي نوع آخر من أنواع الروابط.

يعمل الوب وفق بنية الزبون/مخدّم (Client/Server). فكل مخدّم مسؤول عن مجموعة الوثائق التي لديه فقط، وعندما يتلقى أي طلب من مستخدم من أجل الوصول إلى وثيقة يديرها، يقوم النظام بالبحث عن هذه الوثيقة وإعادتها للمستخدم وفق طريقة ترميز خاصة ومحددة.

يستند الوب على معيارين قياسيين أساسيين هما:

- بروتوكول HTTP (Hypertext Transfer Protocol): وهو يهتم بأمور شبكة الاتصال حتى يؤمن التواصل بين المخدم والزبون (المستخدم) عبر الشبكة.
- لغة التأشير الفائقة HTML (Hypertext Markup Language): وهي تسمح بوصف الوثائق، وخصوصاً الوثائق النصية، لكن يمكن تضمينها أيضاً فبالوسائط الأخرى كالصوت والفيديو والصور. تُعتبر لغة HTML ابنةً للغة أعمّ هي SGML (Structured Generalized Markup Language)، وهي في تطور مستمر منذ بدايتها حتى الآن، وذلك بهدف تحسين طرق التخاطب مع المستخدم وزيادة فعاليتها، ومنذ وقت قريب جداً، أُضيفت تعديلات لجعل صفحات HTML ديناميكية، ولا يزال التطوير مستمراً.

تُحدد بنية الوثائق في هذا النظام باستخدام ارتباطات تشعبية وحيدة الاتجاه ومحفوظة بشكل دائم (ستاتيكية) ضمن الوثيقة نفسها. تؤمن HTML تعليمات خاصة لبناء هذه الارتباطات. عندما يطلب المستخدم وثيقة ما، فإنه يتلقى المعلومات التي ضمن الوثيقة، والارتباطات التشعبية التي مصدرها هذه الوثيقة. يمكن لمفسر HTML (المتصفح Browser) الموجود عند المستخدم أن يتعامل مع الوثيقة، فيُظهر محتواها من المعلومات (من مختلف أنواع الوسائط)، كما يُظهر الارتباطات التشعبية على الشاشة على شكل مناطق حساسة بلون مغاير وتحتها خط (تسمى عادةً Anchors). تُحدد طرق عنونة الوثائق من أجل الوصول إلى وثيقة، عبر ارتباط تشعبي معين، باستخدام طريقة وحيدة وموحدة في العنونة تسمى URL (Unified Resource Locator). كان أول متصفح HTML في العالم هو Mosaic من NCSA، ويتقاسم حالياً البرنامجان MS Internet Explorer, Netscape Navigator سوق المتصفحات في العالم.

يسمح الوب للكاتب وحده بأن يعدّل الصفحات التي يملكها (المسؤول عنها)، ولا يمكن لأحد غيره أن يعدل تلك الصفحات دون إذن منه. لكن بالمقابل، يمكن لأي كاتب أن يشير إلى صفحات أي كاتب آخر، بأن ينشئ رابطة من صفحته إلى تلك الصفحات أينما كان موقعها.

البحث عن المعلومات في الوب

إن طريقة البحث في نظم الوسائط الفائقة هي التصفح، لكن هذه الطريقة تصبح صعبة جداً وغير فعالة مع نظام ضخم جداً مثل الوب. لقد فرضت هذه المشكلة ضرورة استخدام تقنيات نظم استرجاع المعلومات ومكاملتها مع الوب من أجل تحسين عمليات البحث، وبالتالي تسهيل وصول المستخدم إلى طلبه، وكانت النتيجة هي محركات البحث (Search Engines) على الوب.

مع ظهور محركات البحث، أصبح في مقدور المستخدم أن يصوغ طلبات بحث مختلفة، وتنفيذها من قبل محرك البحث، وقد اختلفت طرق صياغة الطلبات وتنوعت، فمن محركات البحث ما سمح باستخدام كلمات مفتاحية في عملية صياغة الطلبات، ومنها ما سمح باستخدام تعابير منتظمة (Regular Expressions) تحدد نموذجاً للكلمات التي نبحث عنها، ومنها ما دعم صياغة الطلبات باستخدام اللغات الطبيعية. وبغية تحسين نتائج البحث، قدمت هذه المحركات تقنيات متنوعة ومتعددة لفلترة النتائج. كما دعمت محركات البحث عمليات البحث بالتجول ضمن الوب، فقد قامت ببناء مجلدات خاصة بها، يحوي المجلد منها على مجموعة من الفئات، وتُوزع المواقع على الوب بين هذه الفئات كل حسب موضوعه، مما يسمح للمستخدم، من خلال عمليات تجول بسيطة، أن يحدد الموضوع الذي يريده، وأن يعثر على المعلومة في وقت أقصر.

لكن التضخم المتزايد للوب أدى إلى تعقيد عملية البحث فيه، فعملية فهرسة وتسجيل صفحات الوب أصبحت عملية طويلة جداً نظراً للزيادة الكبيرة في حجم الوب كل يوم، وهذا ما فرض ضرورة بناء أدوات تساعد في تسجيل الصفحات الجديدة آلياً. وكانت النتيجة بناء أداة مساعدة لعمل محرك البحث سميت عنكبوت محرك البحث Spider، وهو برنامج يقوم بزيارة مواقع الوب، والتجول ضمنها والبحث عن الصفحات الجديدة ليضيفها إلى قاعدة بيانات محرك البحث. يقوم العنكبوت

بزيارة الصفحة الرئيسية في الموقع، والتي تحوي مجموعة من الارتباطات التشعبية إلى بعض صفحات الموقع. يتعرف العنكبوت على مجموعة الصفحات الجديدة هذه، ثم ينتقل إليها ويضيفها إلى قاعدة بيانات محرك البحث المرتبط به. وفي الصفحات الجديدة، يتعرف على الارتباطات التشعبية الموجودة في الصفحة وهكذا... لا يقتصر عمل العنكبوت على زيارة الصفحات الجديدة ضمن الموقع فقط، بل يزور الصفحات القديمة أيضاً ليتأكد من أنها لم تتغير، وهذا من شأنه أن يُصحح نتائج البحث باستخدام محرك البحث.

ومع التطور والمنافسة بين محركات البحث، ظهر مفهوم جديد لمحركات البحث، فالقائمون عليها لم يريدوا أن تبقى محركات البحث محطةً يقصدها المستخدمون عندما يريدون البحث عن معلومات، بل أرادوا أن تكون هدفاً بحد ذاتها، أي أن يرتادها المستخدمون ومنها يقومون بما يشاؤون من عمل. لذا قاموا بإضافة الكثير من الأدوات والخدمات، بعضها مجاني، بغية جذب أكبر عدد من المستخدمين، كما شكّلت مجموعاتٌ من محركات البحث شبكات فيما بينها، تتعاون وتقدم خدمات أكثر تميزاً، ولا تزال المنافسة بين محركات البحث على أشدها. ونظراً إلى أهمية محركات البحث باعتبارها الأدوات الأكثر استخداماً في البحث عن المعلومات في الإنترنت، فقد خصصنا لها أربعة فصولٌ مطولة من هذا الكتاب.

يواجه الوب مشاكل عديدة منها الضياع، وعدم دقة نتائج البحث باستخدام محركات البحث، ومشاكل صيانة قواعد معلومات الوب الضخمة جداً، والتخلص من الارتباطات التشعبية الميتة، ومشاكل الازدحام، و... لكن كل هذه المشاكل على أهميتها، لا تنقص من أهمية الوب، ومركزه كمصدر المعلومات الأكثر غزارة والأيسر وصولاً، ووسيلة التخاطب التي لا غنى عنها في هذا العصر، عصر التغيرات والتطورات السريعة.

فيما تبقى من هذا القسم سنستعرض مجموعة من الأدوات "القديمة" التي كانت مستخدمة بكثافة في عمليات البحث عن المعلومات في الإنترنت قبل انتشار استخدام الوب انتشاراً متنامياً اعتباراً من النصف الثاني من عام 1993. فقد ساهم ظهور الوب ومحركات البحث فيه، والتطور الدائم المستمر

والسريع لهذه المحركات، في أفول نجم هذه الأدوات، وعاماً إثر عام، يتراجع عدد مستخدمي هذه الأدوات تراجعاً كبيراً. وربما لم يسمع إلاّ القليل من مستخدمي الإنترنت الحديثين عن غوفر (Gopher) أو فيرونيكا (Veronica) أو وايس (Wais) أو...

إن هدفنا من عرض هذه الأدوات هو استكمال ما بدأناه سابقاً من شرح شامل لطرق البحث في الإنترنت، وسيكون العرض سريعاً ودون تعمق في التفاصيل مكتفين بطرق البحث وإمكاناته.

ذكرنا بعض العناوين عند شرح طريقة الوصول إلى هذه الأدوات، ويجب التنويه إلى أن هذه العناوين قد لا تكون صالحة، فالكثير من مخدمات هذه الأدوات قد خرجت من العمل، ويأتي ذكرها هنا من أجل شرح طرق الوصول (سابقاً) إلى هذه الأدوات.

4-3-2 غوفر Gopher

نشأ غوفر في جامعة مينيسوتا (Minnesota) عام 1991، على يد Mark P.McCahill وفريق من المبرمجين، وذلك بهدف حل مشكلةٍ كانت تواجههم في الجامعة. فمستخدمو الشبكة من الموظفين كانوا يطلبون دوماً البحث عن نفس المعلومات، مما يؤدي إلى ازدحام الشبكة وتأخر الوصول إلى هذه المعلومات. ومن ثم كان الحل هو في تجميع المعلومات وتصنيفها وتبويبها تبويباً جيداً، بحيث يتمكن المستخدمون من الوصول إلى المعلومة بزمن أقصر، وسهولة أكبر، ودون التسبب بازدحام الشبكة.

لم تحتفظ الجامعة بغوفر لفترة طويلة، إذ سرعان ما انتشر غوفر وأخذ يتعامل مع الشبكة خارج حدود الجامعة. وقد حظي بشعبية كبيرة في تلك الفترة، بحيث أحصي في عام 1995 أكثر من 9000 مخدّم غوفر على الشبكة.

فضاء غوفر

يتألف غوفر من مجموعة من القوائم (Menus)، ترتبط هذه القوائم بعضها ببعض ضمن بنية شبه شجرية، من القوائم ذات المواضيع العامة وصولاً إلى القوائم الأكثر تخصصاً. تشكّل هذه القوائم عملياً

شجرة ضخمة جداً ويجب الوصول إلى وريقات هذه الشجرة التي هي مؤشرات إلى ملفات أو خدمات. القائمة هي مجموعة من العناوين، وكل عنوان هو عنوان لقائمة ابنة للقائمة الحالية، وباختيار عنوان منها، ننتقل إلى القائمة الابنة الموافقة، وهكذا.

يمكن أن تحتوي القائمة على أي نوع من الموارد، مثلاً عنوان إلى مجموعة من الألعاب، وعنوان إلى مكتبة الجامعة، وعنوان إلى الكتاب الذي يحوي أرقام الهواتف، وعناوين إلى أخبار من مجموعات الأخبار، وعناوين إلى وصلات Telnet. كما يمكن الوصول منها إلى أرشيفات FTP باستخدام آرشي (Archie)، والاتصال بـ وايس، كما يمكن الاتصال منها مع مخدمات أخرى لغوفر.

إن مجموعة الموارد التي يمكن الوصول إليها من غوفر تكوّن ما يسمى فضاء غوفر، وهي تتعدى حدود مخدّم غوفر لتشمل جميع موارد الشبكة.

البحث عن المعلومات في غوفر

يجب الانتباه إلى أن غوفر هو أداة بحث بالتصفح، ولا يمكن القيام بصياغة طلبات بحث فيه. كما أن البحث بالتصفح فيه ليس سهلاً وبسيطاً، فهو يعاني من مشاكل كثيرة أهمها أنه مقيد تماماً بالبنية الشجرية مما يؤدي إلى تخاطبية طويلة مع المستثمر غالباً ما تكون محبطة، إضافةً إلى أنه مقيدٌ تماماً بطريقة المؤلف الشخصية في ترتيب المعلومات. لذلك نجد هنا تركيزاً كبيراً على الحاجة وتعريفها تعريفاً دقيقاً جداً، كما يجب تمثّل طريقة المؤلف في التفكير حتى نتمكن من الوصول إلى ما نريد. الطلبات الوحيدة الممكنة فيه هي البحث ضمن القائمة الحالية عن كلمة/جملة ما (مطابقة النماذج Pattern Matching).

إن إمكانيات التصفح بسيطة أيضاً، فلا يمكن الانتقال إلا من قائمة أم إلى قائمة ابنة، أو بالعكس. أما باقي أنواع الانتقال فهي غير ممكنة. ومن أجل دعم هذه التصفح البسيط، ولحل جزء من مشكلة الضياع والتصفح المعقد، يتيح غوفر إمكانية استخدام الإشارات المرجعية (Bookmarks)، ولا يُنصح باستخدام غوفر دون استخدام هذه الإشارات المرجعية أثناء التجوال.

يمكن الوصول إلى غوفر واستخدامه بعدة طرق، منها استخدام برنامج خاص من أجل الوصول إلى مخدمات غوفر، مثل البرنامج WSGopher (لنظام التشغيل ويندوز). كما يمكن الوصول إلى أحد مخدّمات غوفر باستخدام Telnet. نحاول في المثال التالي الدخول إلى مخدم غوفر في جامعة مينيسوتا:

Telnet://consultant.micro.umn.edu

باستخدام كلمة الدخول gopher.

كما يمكن الوصول إلى مخدّمات غوفر من خلال متصفحات الوب، فمن أجل الدخول إلى مخدم غوفر السابق، نستخدم العنوان:

gopher://gopher.micro.umn.edu

4-3-3 فيرونيكا Veronica

إن البحث بالتصفح الذي يسمح به غوفر، ليس فعالاً كفايةً، فعند البحث عن معلومة معينة، يتوجب القيام بعملية تصفح طويلة جداً، تستغرق وقتاً طويلاً، وفي كثير من الأحيان لا نتمكن من الوصول إلى المطلوب. وهنا ظهرت فكرة إنشاء برنامج يقوم بعملية البحث بحثاً آلياً، وعوضاً عن وقت الانتظار الطويل، تُختصر عملية البحث إلى بضع ثوان.

ففي عام 1992، قام S. Foster وF. Barrie من جامعة Nevada بتحقيق هذه الفكرة، فكانت النتيجة محركَ البحث فيرونيكا، الذي يبحث حصراً في قوائم غوفر ليشكلا معاً ثنائياً قوياً للبحث عن المعلومات في الإنترنت حتى وقت قريب.

يمكن البحث باستخدام فيرونيكا عن أي معلومة في فضاء غوفر: صفحات من مختلف الأنواع، أرشيفات USENET، ووصلات Telnet، و... وقد تمكن فيرونيكا من فهرسة معظم مخدّمات غوفر على الإنترنت (إن لم نقل جميعها)، ويقوم بزيارة هذه المواقع مرة كل شهر، ليُعدّل قاعدة بياناته التي تحوي فهارس من قوائم غوفر المختلفة.

البحث عن المعلومات في فيرونيكا

يستند البحث عن المعلومات في فيرونيكا على النموذج البوليناتي، فطلبات البحث في فيرونيكا هي كلمات مفتاحية تربط بينها عوامل بوليانية، لذا فهو غير قادر على حساب درجات المطابقة بين النتائج والطلب. كما أن عملية الفهرسة فيه مقتصرة على عناوين الملفات والصفحات وغيرها من الموارد الموجودة ضمن القوائم، فلا يستطيع الدخول إلى داخل الملفات ليفهرسها، وهذا ما يُضعف من قيمة النتائج التي نحصل عليها.

يُقسم البحث في فيرونيكا إلى قسمين :

أ. *البحث في مجلدات غوفر*: حيث يبحث فيرونيكا في القوائم التي لها قوائم بنات (ليست قوائم أوراقاً في الشجرة الضخمة التي تشكلها قوائم غوفر)، وعندما يجد عنواناً ضمن إحدى هذه القوائم يحتوي على كلمات البحث المستخدمة، يعيده ضمن نتائجه.

ب. *البحث في فضاء غوفر*: وفيه يبحث ضمن القوائم الأوراق. باستخدام هذا النوع من البحث، يمكن البحث عن أي معلومة في فضاء غوفر، من ملفات نصية وملفات صور ومجلدات غوفر وفهرس وايس و...

إن البحث في فيرونيكا غير حساس لحالة الأحرف، كما يمكن (كما ذكرنا سابقاً) استخدام العوامل البوليانية وهي AND, OR, NOT، إضافة إلى إمكانية تجميع هذه العوامل باستخدام الأقواس.

يستطيع المستخدم تحديد مجموعة من الخيارات لطريقة إظهار النتائج والفلترة، وذلك بتمرير وسائط خاصة، فمثلاً البحث tourism –m10 يعني تحديد نتائج البحث بعشر نتائج فقط، والبحث باستخدام الوسيط (t0-) للبحث عن ملفات نصية، و(t1-) للبحث عن مجلدات، و(t8-) للبحث عن وصلات Telnet، وغيرها...

يمكن استخدام فيرونيكا من خلال غوفر، وذلك بفضل الخيار الخاص بذلك والموجود ضمن قوائم غوفر.

4-3-4 البحث في مواقع FTP بواسطة Archie

إن الحاجة لإيجاد مواقع FTP على الإنترنت هي في تزايد مستمر، لذا ظهرت مواقع على الوب تهتم بتصنيف هذه المواقع على الإنترنت، وتعرضها للمستخدم الذي يقوم بالبحث بالتصفح ضمنها. وقد زوّدت بعض المواقع الزوار بأداة للبحث المحلي، فبات بإمكان المتصفح لموقع FTP أن يبحث فيه عن ملفات يطلبها. إلا أن هذه الأدوات المحلية لم تفِ بالغرض تماماً، فكان لا بد من أداة بحث أكثر شمولية، وهذا ما أدى إلى انتشار استخدام الأداة آرشي، وهي أداة للبحث عن مواقع FTP على الإنترنت.

قبل الاستمرار في الحديث عن آرشي وآلية عمله، نحتاج لإعطاء لمحة صغيرة عن مواقع FTP.

لمحة عن مواقع FTP

البروتوكول FTP هو بروتوكول لنقل الملفات على الإنترنت من موقع لآخر. تُعطي مواقع FTP لكل مستخدم رمز دخول وكلمة سر خاصّين به، وذلك حتى يتمكن من الولوج إلى الموقع والاستفادة من الملفات المتاحة عليه. يتيح الكثير من هذه المواقع إمكانية الولوج لأي شخص يزور الموقع، وذلك بتوفير كلمة دخول مشتركة (ودون كلمة سر) لكل من يريد الولوج إلى الموقع. كلمة الدخول هذه هي anonymous، وفي الموقع يجد الزائر مجموعة من الملفات التي يسمح مدير الموقع بالحصول عليها، بينما تبقى الملفات الخاصة في الموقع بعيدة عن متناول مستخدمي anonymous. وإذا أراد المستخدم الحصول على ملفات أكثر ضمن الموقع، فيجب أن يكون لديه رمز دخول وكلمة سر خاصّين به.

آلية عمل Archie

طُوِّر آرشي (Archie) في جامعة McGill على يد Alan Emtage وزملائه. وقد بدؤوا في أول الأمر بزيارة مواقع FTP على الإنترنت باستخدام رمز الدخول anonymous، وفهرسة الملفات الموجودة ضمنه يدوياً. ثم تطور العمل إلى عمل آلي تماماً يقوم به مخدم خاص بآرشي. وفي وقت لاحق، تحول المشروع إلى مشروع تجاري بمساعدة شركة خاصة، وأصبح بمقدور آرشي أن يفهرس عدداً أكبر من مواقع FTP، وأن يصل إلى جميع الملفات ضمن هذه المواقع (بعد إذن مديري هذه المواقع طبعاً)، وفهرستها كاملة. يوجد آرشي على مخدمات خاصة به، تزور مواقع FTP التي تعرفها كل شهر مرتين، لتلاحق أي تغييرات على هذه المواقع.

يمكن استخدام آرشي بعدة طرق منها استخدام برنامج خاص بذلك مثل WSaArchie (لنظام التشغيل ويندوز)، أو الوصول إلى أحد مخدمات آرشي باستخدام Telnet:

Telnet://archie.sura.net

واستخدام رمز الدخول archie. كما يمكن استخدام E-mail، حيث يكتب المستخدم أوامر البحث ضمن رسالته، ويوجهها إلى بريد أحد مخدمات آرشي مثل:

archie@archie.sura.net

البحث في آرشي

إن البحث في آرشي هو بحث من نمط البحث عن الوصفات، كما أنه يعاني من تعقيد واجهات الاستخدام. يمكن البحث ضمن مخدم آرشي باستخدام الأمر find، يليه طلب البحث. مثلاً:

find geography سيبحث عن جميع الملفات (على جميع مواقع FTP التي يعرفها) التي تحتوي ضمن أسمائها كلمة geography.

يتيح آرشي مجموعة كبيرة من خيارات البحث، فيمكن استخدام رموز عامة متنوعة جداً في صياغة الطلبات، فالبحث عن ^geogra يعني البحث عن جميع الملفات التي تحوي geogra في أول اسمها.

أما graphy$ فتعني البحث عن الملفات التي ينتهي اسمها بـ graphy. ويمكن استخدام غير ذلك من الرموز المستخدمة في التعابير المنتظمة (Regular Expressions) في نظام UNIX.

يتيح آرشي أيضاً مجموعة كبيرة من الإعدادات، منها إعدادات خاصة بطريقة ترتيب النتائج وإظهارها، وطريقة معاملة الطلب (مراعاة حالة الأحرف مثلاً...)، أو تحديد Domain Name معين للمواقع التي نريد البحث ضمنها، أو حتى المجلدات التي سيبحث فيها فقط.

4-3-5 البحث في النصوص بواسطة WAIS

نأتي الآن إلى أداة أخرى يمكن اعتبارها أول نظام استرجاع معلومات (IRS) حقيقي على الشبكة، وهي وايس (Wide Spread Information Server) WAIS. لقد رأينا كيف أن غوفر وفيرونيكا يسمحان بالبحث عن ملفات من مختلف الأنواع، لكن هذا البحث هو بحث مقتصر على فهرسة العناوين فقط، دون الخوض في محتوى الملفات، وهذا ما جعلهما قاصرين عن تلبية حاجات المستخدمين على نحوٍ فعال. إن الحاجة إلى نظام استرجاع معلومات هي التي دفعت Brewster Kahle بدعم من مجموعة من الشركات إلى إنشاء النظام وايس. وبعد فترة من إنشائه، اشترته شركة America On Line (AOL)، مما يدل على الأهمية الكبيرة له كنظام بحث عن المعلومات على الإنترنت.

البحث عن المعلومات في وايس

يبحث وايس في محتوى الوثائق باستخدام الكلمات المفتاحية والعوامل البوليانية بينها. يستند في ذلك على نموذج بولياني موسع (Extended Boolean Model). إضافةً إلى ذلك يستطيع إيجاد أمكنة الكلمات المفتاحية داخل النصوص، ويسمح بترتيب الوثائق الأجوبة حسب درجة المواءمة المتناقصة. كما يسمح أيضاً بالتغذية الراجعة Relevance Feedback لأجل تحسين النتائج وذلك إما بتعديل الطلبات وإما باستخدام إحدى الوثائق الأجوبة. مشاكله الأساسية تكمن في أنه يبحث في النصوص

فقط وفي كون واجهاته معقدة الاستخدام. تسعى مخدمات وايس إلى التخصص بموضـوع معين، فعلى العكس من غوفـر، توجد مخدمات خاصـة لـلأدب، وأخرى خاصـة بالأخبار، وغيرها بمواضيع الرياضيات فقط و...

تتميز هذه المخدّمات بأن نهاية أسمائها هي src، فمثلاً المخدّم الخاص بوثائق الإنترنت المعيارية هو: internet-standards.src.

يمكن استخدام وايس بطرق عدة، منها استخدام برنامج خاص مثل WinWai، أو استخدام Telnet:

Telnet://wais.wais.com

أو عن طريق مستعرضات الوب، من خلال العنوان:

http://www.wais.com/newhomepage/wais-dbs.html

أما طريقة الاستخدام فسهلة جداً، إذ يكفي اختيار قـاعدة البيانات التي نريد البحث ضمنها (أو مخدم وايس الذي يحويها)، ثم كتابة كلمات البحث، وضغط زر البحث. الفرق الوحيد بين استخدام برنامج خاص لاستخدام وايس وبين استخدام الوب هو عدم إمكانية تحديد أكثر مـن قـاعدة بيانـات واحدة للبحث ضمنها عند استخدام الوب، بينما يمكـن تحديـد عـدة قواعـد بيانـات عنـد استخدام برنامج خاص.

الفصل الخامس

محركات البحث وطرق تقييمها

لا شك بأن أهم أدوات البحث عن المعلومات في الإنترنت هي محركات البحث (Search Engines)، لذلك خصصنا لها جزءاً هاماً من هذا الكتاب. وباعتبار أنه توجد المئات من محركات البحث في الإنترنت، كان لا بد لنا من اختيار أهم هذه المحركات وأكثرها استخداماً ودراستها دراسة عميقة بهدف الإحاطة بجميع مزاياها والخدمات التي تقدمها من ناحية، ومقارنتها وفق مجموعة من المعايير بهدف تحديد أفضل المحركات، كي يستطيع القارئ الكريم من أجل كل طلب معين من المعلومات تحديد محرك البحث الأنسب لتحقيق هذا الطلب.

لذلك نعرّف في هذا الفصل بمحركات البحث وآلية عملها، ونحدد مجموعة من المعايير المناسبة لمقارنة المحركات بعضها ببعض.

5-1 تعريف محركات البحث

محرك البحث (Search Engine) هو البرنامج الذي يسمح للمستخدمين بالبحث ضمن الوثائق الموجودة في موقع وب أو مجموعة مواقع وب. تحتوي محركات البحث البسيطة على بعض فئات المواضيع التي يمكن البحث فيها بواسطة المطابقة المباشرة. فمثلاً، قد يحتوي موقع وب لبيع السيارات المستعملة على محرك بحث يسمح بالبحث عن مصنعي السيارات ويعيد النتائج على شكل

جداول تحتوي الماركة والنموذج. في حين تسمح محركات البحث المعقدة باستخدام الكلمات المفتاحية في كتابة الطلبات، ومن ثم تكون نتائج البحث فيها هي الوثائق التي تحتوي الكلمات المفتاحية المحددة في الطلب. يُعتبر البحث بالكلمات المفتاحية، من وجهة نظر المستخدم، أكثر فاعلية لأنه يسمح للمستخدم بتعريف الطلب الذي يناسبه دون أن يعتمد على طرق تصنيفية سابقة التعريف في محرك البحث. لكن، وفي الوقت نفسه، يُعتبر هذا النمط من محركات البحث صعب البرمجة والتنفيذ من وجهة نظر المبرمجين، إضافةً إلى كونه بطيء التنفيذ.

عموماً، تُعتبر عمليتي البحث والفرز (Searching & Sorting) من أكثر الأعمال التي يقوم بها الحاسوب تطلباً للحساب الكثيف، ولا يكاد يخلو أي منهاج في علوم الحاسوب من مواد مفصّلة في هذين الموضوعين. ويجب على جميع محركات البحث التي تحتوي المواصفات الأساسية أنْ تتضمن كلتا العمليتين. إذ، كما رأينا في الجزء الأول من هذا الكتاب، أي طلب من محرك البحث يحوّل إلى سلسلة محارف يجري البحث عنها في الوثائق، أو في فهارس الوثائق، وتُقيِّم نتيجة كل بحث، ثم تُرتب (تُفرز) الوثائق الأجوبة حسب درجات المطابقة المتناقصة. آخذين بعين الاعتبار سرعة استجابة محركات البحث، يجب أن يكون رماز (Code) برنامج محرك البحث مقتصداً (Optimized) قدر الإمكان وسريع التنفيذ، فنادراً ما تُستخدم محركات البحث البطيئة.

يجري وضع وتنفيذ محرك البحث على مخدّم الموقع، كما يجب أيضاً وضع الوثائق أو فهارس الوثائق، التي يمكن تخزينها في قواعد بيانات مخصصة، على مخدّم الموقع. ومن الضروري أن يكون هذا المخدّم حاسوباً متفوقاً (بسبب كلفة تنفيذ عمليات البحث والفرز)، وفي حالة محركات البحث التي نجدها على الإنترنت، والتي قد تتضمن فهارسَ لملايين المواقع، يمكن تخصيص مخدّمات بكامل طاقتها لتنفيذ برنامج محرك البحث، خصوصاً أنها قد تضطر لمعالجة آلاف طلبات البحث في آن واحد.

إذا كان الموقع بسيطاً فيمكن برمجة محرك بحث فيه بسهولة. أمّا في حالة محرك بحث متطور يدعم البحث بالكلمات المفتاحية، ويسمح ببعض خواص نظم استرجاع المعلومات مثل استخدام العوامل البوليانية أو النموذج الشعاعي، فيمكن اللجوء إلى بعض البرمجيات الجاهزة. نذكر مثلاً برنامج

Excite for Web Server الموجود في الموقع: /http://www.excite.com/navigate، والذي يسمح بكتابة طلبات معقدة نسبياً.

سوف نترك كيفية برمجة محركات البحث إلى كتب متخصصة أخرى ضمن هذه السلسلة، في حين نركز الاهتمام في الفصول القادمة من هذا الكتاب لدراسة محركات البحث الشائعة الاستخدام والمتعددة الأغراض المتوفرة بكثرة في الإنترنت.

5-2 معايير تقييم محركات البحث

لقد اخترنا مجموعة من المعايير لمقارنة محركات البحث استقيناها من خبرتنا الخاصة في مجال البحث عن المعلومات، إضافةً إلى مجموعة أخرى من المعايير كنا قد رصدناها أثناء التحضير لهذا الكتاب في مجموعة كبيرة من المواقع المتخصصة في هذا الموضوع. ترتبط هذه المعايير بآليات البحث عن المعلومات (بنوعيه المتقدم والبسيط) التي كنا قد غطينا خلفياتها النظرية تغطيةً محدودة في الجزء الأول من الكتاب، بفاعليتها وطرق فلترة النتائج وإظهارها والسرعة في التنفيذ وحجم قاعدة البيانات التي يستخدمها محرك البحث.

5-2-1 البحث البسيط

المقصود بالبحث البسيط هو عمليات البحث التي تعتمد على مطابقة النماذج (Pattern Matching) من أجل إيجاد الوثائق المطابقة للطلب (الاستعلام).

يمكن إيضاح مطابقة النماذج بالمثال البسيط التالي:

إن كلمتي وثيقة ومستند لهما نفس المعنى، لكن عند القيام بالبحث البسيط عن كلمة "مستند"، فيجب أن توجد كلمة "مستند" ضمن الصفحة حتى يعتبر النظام هذه الصفحة مقبولة. أما إذا وجدت

صفحة تحوي كلمة "وثيقة"، فلن يتم اعتبارها مقبولة. إن أبسط مثال عن البحث البسيط هو عمليات البحث التي تؤمنها محررات النصوص المختلفة (مثل MS Word أو المفكرة Note Pad).

تسمح جميع محركات البحث بعمليات البحث البسيط، وسنقارن بين عمليات البحث البسيط هذه حسب المعايير التالية:

آ. التحسس للأحرف الصغيرة والكبيرة

هذا المعيار لا يمكن تطبيقه على محركات البحث العربية (عند كتابة طلبات باللغة العربية) لأننا في اللغة العربية نتعامل مع الأحرف بطريقة واحدة، أما في اللغة الإنكليزية (وغيرها) فيمكن كتابة الحرف بطريقتين: صغير أو كبير.

إن سماح محرك البحث للمستخدم باستخدام الأحرف الكبيرة أو الصغيرة دون تمييز في كتابة طلباته يمنح المستخدم حرية ومرونة أكبر عند صياغة الطلبات، فهو لن يحتاج إلى التأكد من حالة لوحة المفاتيح (زر Caps Lock) في كل مرة يصوغ فيها طلباً، أو إلى أي إشارة أخرى عن حالة الأحرف ضمن النصوص التي يبحث فيها. بالمقابل قد تمنح إمكانية البحث البسيط مع مراعاة حالة الأحرف المستخدمَ إمكانيةً أكبر في تحديد المعلومات التي يريدها، كما في حالة البحث عن الوثائق التي تحوي كلمة معينة شريطة أن تكون هذه الكلمة في أول الفقرة (وهنا يجب أن يكون الحرف الأول كبيراً). إن إعطاء المستخدم إمكانية تحديد طريقة البحث البسيط (مراعاة حالة الأحرف أم لا) قد تكون الحل الأسلم، وهذا ما نلاحظه في بعض محررات النصوص التي تسمح للمستخدم أن يحدد إن كان يريد البحث مع مراعاة حالة الأحرف أم لا (كما في NotePad وتحدد بالخيار Case Sensitive).

ب. تجميع كلمات البحث في جمل

في بعض الأحيان، يحتاج المستخدم للبحث عن الصفحات التي تحوي جملة بكاملها أو مجموعة من الكلمات معاً (مثل: "بلاد الشام" أو "المجموعة الشمسية")، إن البحث عن كلمة "بلاد" ثم عن كلمة

"الشام" لن يعطي النتيجة المرادة، فالمستخدم يريد الصفحات التي تحوي "بلاد الشام" وليس "بلاد" و"الشام".

إن إتاحة هذه الميزة قد تكون مفيدةً للمستخدم من أجل صياغة طلباته، لكن لا تتيح جميع المحركات هذه الميزة، لذا يجب على المستخدم معرفة هذه المعلومة عن محرك البحث الذي يستخدمه، فقد يعطي نتائج مخالفة لما يتوقعه.

ج. الكلمات المفتاحية اللازمة

يحتاج المستخدم في بعض الأحيان إلى أن تحوي الصفحات الناتجة عن البحث كلمة معينة ذات أهمية خاصة بالنسبة له. مثلاً، إذا أراد المستخدم البحث عن الصفحات التي تتحدث عن السياحة في سورية، وعن وسائط النقل في سورية أيضاً، فإن كلمة "سورية" هي كلمة أساسية في البحث الذي يطلبه، وبالتالي يريد أن توجد ضمن جميع الصفحات الناتجة. أما كلمة "سياحة" أو كلمة "نقل" فقد توجد أو لا توجد ضمن النتائج، فوجود إحدى هاتين الكلمتين (إضافة إلى وجود "سورية") مطابق لطلب المستخدم.

د. الكلمات المفتاحية الممنوعة

قد يحتاج المستخدم في بعض الأحيان إلى صياغة طلبات ذات طابع محدد. لنفترض مثلاً أن المستخدم يريد البحث عن الصفحات التي تتحدث عن الوسائط المتعددة لكنه لا يريد الصفحات التي تتحدث عن الصوت، عندها يحتاج إلى طريقة تعبير معينة يجب أن يؤمنها له محرك البحث من أجل استثناء الصفحات التي تتحدث عن الصوت، وتسمى الكلمة المفتاحية "الصوت" في هذه الحالة الكلمة المفتاحية الممنوعة.

هـ. الرموز العامة

قد يلزم المستخدم في بعض الأحيان البحث عن كلمات لها بنية معينة، مثلاً الكلمات التي تبدأ بحرف "أ"، أو الكلمات التي تبدأ بـ "حر"، أو تحتوي مجموعة أحرف أو تنتهي بمجموعة أحرف.

تؤمن بعض المحركات هذه الإمكانية، لكن تجب معرفة طريقة الصياغة، فمثلاً يمكن أن يُستخدم الرمز * كما في *Hyper الذي يشير إلى الكلمات التي تبدأ بالمقطع Hyper.

و. ترتيب الأجوبة

بعد الحصول على مجموعة الصفحات المطابقة لطلب المستخدم، يقوم النظام بعرضها للمستخدم الذي يقوم بتصفحها باحثاً عن المعلومات التي يريدها. لكن عندما يكون عدد الصفحات التي يعيدها النظام كبيراً جداً (وهو الحال في معظم عمليات البحث)، فإن من المستحيل على المستخدم أن يتصفحها جميعاً، وهنا يأتي دور محرك البحث في القيام بعملية ترتيب لهذه الوثائق حسب درجات مطابقتها لطلب المستخدم.

لا تقوم جميع محركات البحث بعملية الترتيب هذه، وتكمن المشكلة في تابع المطابقة المستخدم، فبعض توابع المطابقة لا تستطيع حساب درجات مطابقة الصفحات لطلب المستخدم.

5-2-2 البحث المتقدم

الميزة في هذا البحث هي إمكانية صياغة طلبات يتم فيها تحديد علاقات بين الكلمات المفتاحية المستخدمة. ففي البحث البسيط وجدنا أن البحث يجري على أساس كلمات دون أن نستطيع تحديد علاقات بين هذه الكلمات (عدا الحالة التي يمكن فيها استخدام جمل في عمليات البحث، ففي هذه الحالة نحدد أن العلاقة بين الكلمات ضمن الجملة هي علاقة تجاور)، أما في البحث المتقدم فيمكن أن نحدد الطلب تحديداً دقيقاً، بمجموعة من العلاقات بين الكلمات المستخدمة. نظرياً، نستطيع تحديد أي علاقة بين هذه الكلمات، لكن عملياً تكون العلاقات المتاحة هي علاقات منطقية

(بوليانية)، أو علاقات بسيطة تعتمد على مواضع ورود الكلمات في النصوص. سنناقش هذه العلاقات فيما يلي.

آ. استخدام العوامل البوليانية

لنفترض أن طلب المستخدم يأخذ الشكل التالي:

"الصفحات التي تتحدث عن معارك الحرب العالمية الثانية بين فرنسا وألمانيا أو بين ألمانيا وإنكلترا".

إن البحث البسيط لا يمكن أن يسمح لنا من خلال بحث واحد بالحصول على الصفحات المطلوبة. ففي البحث البسيط، سنقوم بالبحث عن صفحات المعارك بين فرنسا وألمانيا، ثم صفحات المعارك بين ألمانيا وإنكلترا، بعد ذلك، نبحث في المجموعتين عن المعلومات التي نريدها. إن استخدام العاملين البوليانيين "و AND" و "أو OR" يسمحان لنا بصياغة الطلب بشكل أبسط وأسهل، فيمكن أن يصبح الطلب السابق: [معارك الحرب العالمية الثانية "و" ((فرنسا "و" ألمانيا) "أو" (ألمانيا "و" إنكلترا))].

يمكن أيضا استخدام معامل ثالث هو "النفي NOT" في صياغة الطلبات.

ب. البحث المقارب

يستخدم البحث المقارب من أجل تحديد علاقة موضعية بين الكلمات المفتاحية ضمن الطلب. فمثلا يصبح بالإمكان صياغة طلبات من نمط: "الصفحات التي تحوي الكلمتين "بنية" و "حاسوب" على أن لا تكون المسافة بين الكلمتين أكثر من ثلاث كلمات".

5-2-3 الفلترة

يوجد لدينا عدة أنواع من الفلترة، فقد تكون جغرافية مثل: "الصفحات التي تتحدث السياحة في سورية فقط"، أو تاريخية مثل: "الصفحات التي تتحدث عن الذكاء الصنعي، والمنشورة قبل عام

1996"، أو مرتكزةً على الموضوع أو أنواع الوسائط أو العناوين أو أنماط الملفات أو URL أو الارتباطات التشعبية.

5-2-4 تصنيف المواضيع

تتيح بعض محركات البحث إمكانية تصنيف المواضيع في فئات متخصصة، فيمكن للمستخدم أن يحدد الصفحات التي يريدها حسب موضوع معين. تصنف مثلاً المواضيع المتعلقة بالمعلوميات في فئة خاصة، في حين تصنف مواضيع الثقافة والتعليم في فئة أخرى، وهكذا ...

يسمح هذا التصنيف بتحديد البحث تحديداً أفضل مما يعطي دقة أعلى للنتائج، فمن دون تحديد الفئة المتخصصة، يمكن أن تحصل على وثائق (إجابة) لا علاقة لها بالموضوع الذي نريده.

5-2-5 التغذية الراجعة

وتعني أن يعيد المستخدم، بعد قيامه بعملية بحث عن معلومة ما واطلاعه على بعض النتائج التي حصل عليها، صياغة الطلب حسب المعلومات الجديدة التي حصل عليها.

لا تسمح جميع محركات البحث بهذه العملية التي يمكن أن تكون مفيدة جداً ولها تأثير كبير على دقة نتائج البحث.

5-2-6 التحكم بالإظهار

من المهم جداً أن يتمكن المستخدم من القيام بعمليات البحث بسهولة، ويجب على محرك البحث أن يؤمن واجهات تخاطب واضحة وأليفة. كما يجب أن يتيح محرك البحث للمستخدم بالتحكم بكيفية عرض نتائج البحث.

5-2-7 السرعة

من المهم جداً عند القيام بعمليات بحث أن يحصل المستخدم على نتائج بأسرع وقت ممكن، فالمستخدم ينتظر أمام شاشة الحاسوب، ولا يجب أن يطول انتظاره. كلما ازدادت سرعة إنجاز عملية البحث، كلما كان محرك البحث أفضل (طبعاً مع مراعاة كون النتائج جيدة)، لذا ستكون السرعة عاملاً أساسياً في تقييم محرك البحث.

5-2-8 حجم قاعدة البيانات

ترتبط بكل محرك بحث قاعدة بيانات خاصة به. تحتوي هذه القاعدة معلومات عن الصفحات التي يستطيع الوصول إليها، وعناوينها. وكلما ازداد حجم قاعدة البيانات هذه، كلما تمكن محرك البحث من الوصول إلى صفحات أكثر. إن ازدياد حجم قاعدة البيانات هو أمر محبّذ، لكن يجب أن ترافقه جودة في تصنيف نتائج البحث، لأن نتائج البحث ستكون كثيرة جداً من أجل معظم عمليات البحث، ومن ثم لن يتمكن المستخدم من الوصول إلى المعلومات التي يريدها بسهولة. كما يجب أن ترافقه أيضاً سرعة في استجابة محرك البحث.

5-2-9 عوامل أخرى

يمكن أيضاً المقارنة بين محركات البحث من حيث كونها برمجيات خدمية تقدم خدمات متنوعة للمستخدمين. وفي هذه الحالة تكون المعايير غير مرتبطة بعمل وآلية البحث عن المعلومات، وإنما بالمزايا التي تقدمها هذه البرمجيات. نذكر من هذه المعايير مثلاً:

- استخدام تصنيف للمواضيع بمجلدات.
- طريقة تسجيل المواقع ضمن محركات البحث (أو بعبارة أدق ضمن قواعد البيانات الخاصة بهذه المحركات)، والزمن الذي تستغرقه عملية التسجيل هذه.

* طريقة فهرسة محرك البحث للموقع المضاف (يدوية أم آلية).

* .شعبية محركات البحث وأهميتها ، وزمن استخدامها في البحث أو في الانطلاق إلى مواقع أخرى.

نستعرض في الفصل السادس أهم محركات البحث العامة، ثم نستعرض في الفصل السابع بعضاً من محركات البحث البينية والعربية. في الفصل الثامن نقارن بين محركات البحث العامة وفقاً للمعايير التي حددناها في هذا الفصل.

الفصل السادس

محركات البحث العامة

6–1 محرك البحث !Yahoo

www.yahoo.com

يُعد المحركُ الأكثر شعبية، ويعتبر أهم نقطة بداية للبحث، حيث يمكن البحث فيه عن أي شيء. تكمن قوته الأساسية في ترتيب وتجميع المعلومات حسب موضوعاتها ضمن فئات (Categories) مرتبة بشكل شبه شجري[1]، بحيث تكون الفئات ذات المواضيع العمومية في العقد العليا من الشجرة، بينما تكون الفئات ذات المواضيع المتخصصة قريبة من قاع الشجرة. تحتوي كل فئة على مجموعة من الوثائق ومجموعة من الفئات التخصصية المتفرعة عنها، كما يمكن أن تنتمي الفئة إلى عدة فئات أعمَّ منها. فالفئة الخاصة بسوريا مثلاً تنتمي إلى عدة فئات أعمّ منها كما يُظهر الشكل (6–1)، ويمكن

[1] نقول بنية شبه شجرية لأن للشجرة مفهوماً خاصاً في علوم الحاسوب.

تمثيلها بشكل شبه شجري كما يُوضّح الشكل (6-2). لاحظ أن جهة المتراجحة تشير إلى الفئة المتخصصة أكثر، ويمكن الانتقال من فئة إلى فئة باستخدام الارتباطات التشعبية الموجودة في كل فئة.

يحتوي Yahoo! على نصف مليون موقع تقريباً، مقسمة إلى حوالي 25.000 فئة، حيث يقوم مصممو هذه المواقع بفهرستها يدوياً وتحديد الفئات التي تنتمي لها، ثم الاتصال بـ Yahoo! (عن طريق تعليمات معينة سنشرحها لاحقاً) حتى يتمكن من إضافتها إلى قاعدة بياناته في مكانها المناسب.

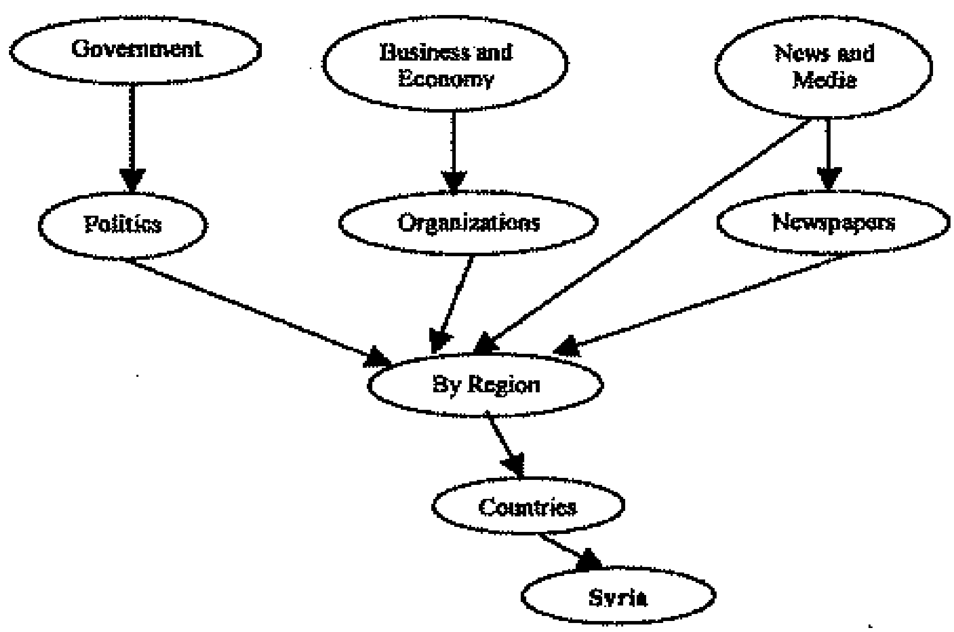
Government > Politics > By Region > Countries > Syria

Business and Economy > Organizations > By Region > Countries > Syria

News and Media > Newspapers > By Region > Countries > Syria

News and Media > By Region > Countries > Syria

الشكل (6-1): بعض الفئات التخصصية التي تنتمي إليها الفئة الخاصة بسوريا

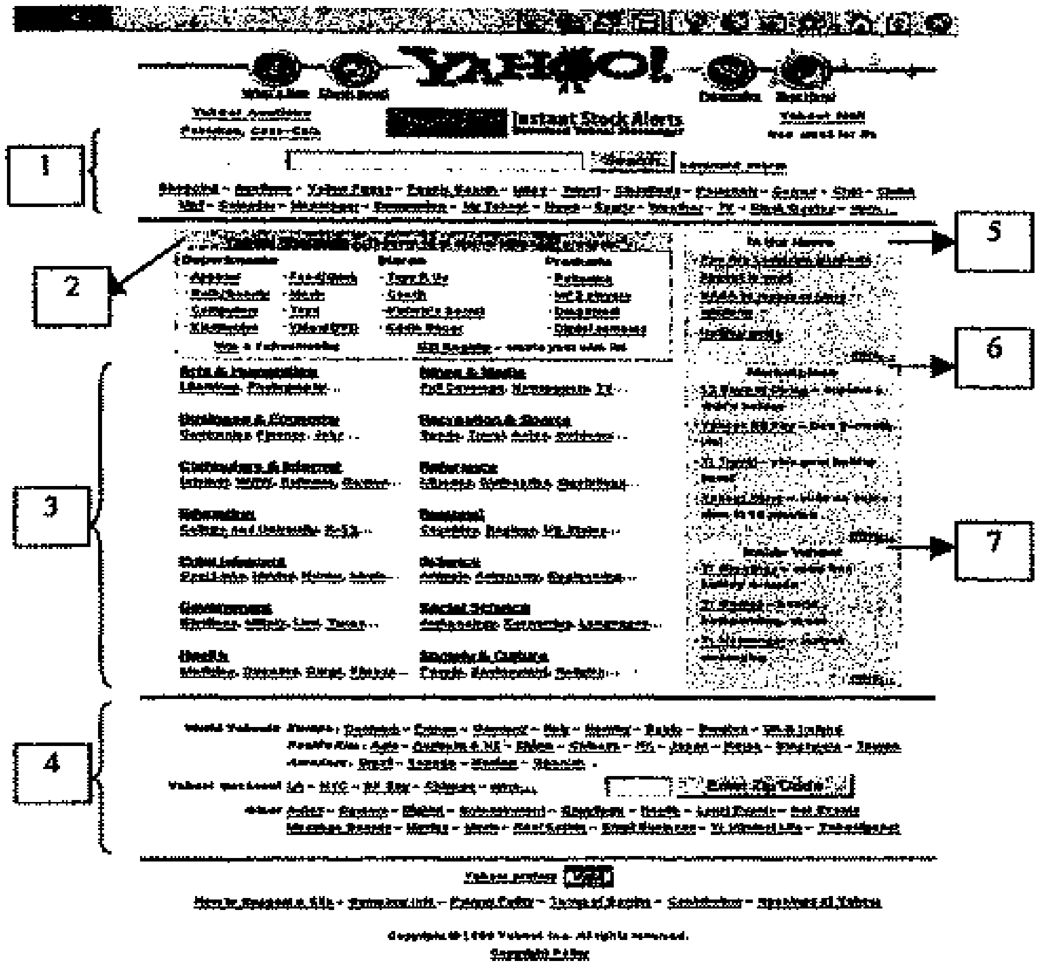


الشكل (6-2): التمثيل شبه الشجري لفئات Yahoo!

يُنصح باستخدام Yahoo! في الحالات التي يكون فيها لدى المستخدم فكرة عامة عن الموضوع، لأن Yahoo! يقدّم أفضل النتائج عندما تُستخدم الكلمات المفتاحية العامة في الطلبات. فمثلاً عند البحث عن مسرحية ما، نبحث أولاً عن الكلمة المفتاحية "مسرحية"، عندئذٍ يعيد Yahoo! الفئة المخصصة للمسرحيات، والاحتمال كبير أن تكون المسرحية التي نبحث عنها ضمن هذه الفئة. ونذكر أنه في حال كان المستخدم على علم تام بما يريد، فليقم بالبحث عن الموضوع حسب العنوان.

6-1-1 واجهة Yahoo!

يُظهر الشكل (6-3) صفحة Yahoo! على الوب التي نصل إليها باستخدام العنوان http://www.yahoo.com. تحتوي هذه الصفحة جميع إمكانات البحث التي يقدمها yahoo! والعديد من خدمات المعلومات الأخرى. نجد فيما يلي الإمكانات المختلفة لـ Yahoo!:



الشكل (6-3): صفحة !Yahoo

* الجزء [1]: يسمح للمستخدم بالقيام بعمليات بحث.
* الجزء [2]: يسمح للمستخدم بالقيام بعمليات تسوق.
* الجزء [3]: يسمح للمستخدم بالتجول ضمن فئات !yahoo.

- الجزء 4: يحتوي مجموعة من الارتباطات التشعبية التي تسمح بزيارة مواقع Yahoo! في دول مختلفة من العالم، أو في مناطق داخل أمريكا. كما نجد مجموعة من الارتباطات التي تسمح بالوصول إلى أدوات وخدمات أخرى من Yahoo!.
- الجزء 5: يسمح للمستخدم بالوصول إلى الأخبار اليومية.
- الجزء 6: يوفر مكاناً لتسويق بعض خدمات Yahoo!.
- الجزء 7: يقدّم بعض الخدمات المجانية الأخرى لـ Yahoo!.

سنعود إلى شرح معظم هذه الإمكانات لاحقاً، وخصوصاً آليات البحث عن المعلومات وهي اهتمامنا الأساسي في هذا الكتاب.

6-1-2 البحث عن المعلومات في Yahoo!

يتيح Yahoo! الأنواع الثلاثة من طرق البحث عن المعلومات: التصفح، البحث عن المحتوى، البحث عن الوصفات.

آ. التصفح

لا تختلف صفحة Yahoo! في شيء عن بقية صفحات الوب، إذ توجد فيها ارتباطات تشعبية إلى مواقع متعددة ومتنوعة وخدمات وأخبار ...

لكن تجدر الإشارة إلى وجود نوع خاص من التصفح وهو الانتقال بالتخصيص أو بالتعميم من فئة إلى أخرى. يحتوي الجزء 3 الموضّح في الشكل (6-3) الفئات الأساسية المعرّفة في Yahoo!، مثل:

Arts & Humanities, Business & Economy, Computer & Internet, ..., Society & Culture.

ومنها ينتقل المستخدم بالتخصيص وصولاً إلى الفئة التي يبحث عنها أو التي يمكن أن تحتوي حاجته من الوثائق أو المواقع.

ملاحظة

يمكن للمستخدم أن يلاحظ وجود العلامة @ ضمن عناوين بعض الفئات المتخصصة. تدل هذه العلامة على أن هذه الفئة مدرجة ضمن أكثر من فئة أكثر عمومية. فمثلاً فئة "Fashion" مدرجة ضمن فئة "Arts" وأيضاً ضمن فئة "Society and Culture"، لذا نجد أن العلامة @ تسبق اسم الفئة Fashion.

ب. البحث عن المحتوى

يُتيح Yahoo! كتابة طلبات (على غرار نظم استرجاع المعلومات) وذلك بأن يقوم المستخدم بكتابة مجموعة من الكلمات (بينها فراغات) في الحقل الحواري الموجود في الجزء [أ]، ثم ينقر الزر Search الموجود على يمين الحقل. كما يمكن للمستخدم أن يجري بحثاً متقدماً، من خلال الارتباط التشعبي Advanced Search (سنتحدث عنه لاحقاً).

يقوم Yahoo! بعدئذٍ بالبحث ضمن قاعدة بياناته التي تتألف من خمسة أقسام هي:

- Categories : فئات Yahoo!.
- Web Sites : مواقع Yahoo!.
- Web Pages : صفحات الوب.
- Related News : الأخبار.
- Net Events : أحداث الشبكة.

يقوم Yahoo! بالبحث أولاً ضمن الفئات، ويعيد النتائج التي يجدها إلى المستخدم، ثم يبحث ضمن المواقع ويعيد النتائج ضمن صفحة مختلفة خاصة بالمواقع. إن لم يستطع العثور على أية صفحات مطابقة، ينقل الطلب إلى محرك بحث مساعد هو Inktomi الذي يقوم بعملية بحث على كامل الشبكة. يمكن التنقل بين نتائج البحث حسب المناطق، من فئات Yahoo! إلى مواقع Yahoo! إلى

الأخبار وغيرها، باستخدام الارتباطات التشعبية التي تظهر أعلى وأسفل صفحة النتائج ضمن جدول من سطر واحد.

محرك البحث الرافق لـ Yahoo! والمسمى Inktomi هو محرك بحث آلي على نحو كامل، فالفهرسة آلية، كما أن عملية العثور على الصفحات آلية أيضاً. إذ يقوم Inktomi بزيارة المواقع على الشبكة بانتظام، باحثاً عن مواقع جديدة أو عن صفحات جديدة، وعندما يجدها، يقوم بفهرستها آلياً، ثم يحفظ الكلمات المفتاحية الخاصة بها ضمن قاعدة بياناته.

الجدير بالذكر أن Inktomi لا يعمل إلا إذا طلب منه Yahoo! ذلك، أي في الحالة التي لا يتمكن فيها Yahoo! من إيجاد أي صفحة مطابقة للطلب في قاعدة بياناته.

يمكن التمييز بين النتائج التي يعطيها Yahoo! من فهرسه اليدوي الخاص، وتلك التي يعطيها Inktomi، حيث تظهر نتائج Yahoo! ضمن المنطقة Web Sites (مواقع Yahoo!)، أمّا نتائج Inktomi فتظهر ضمن المنطقة Web Pages (صفحات الوب).

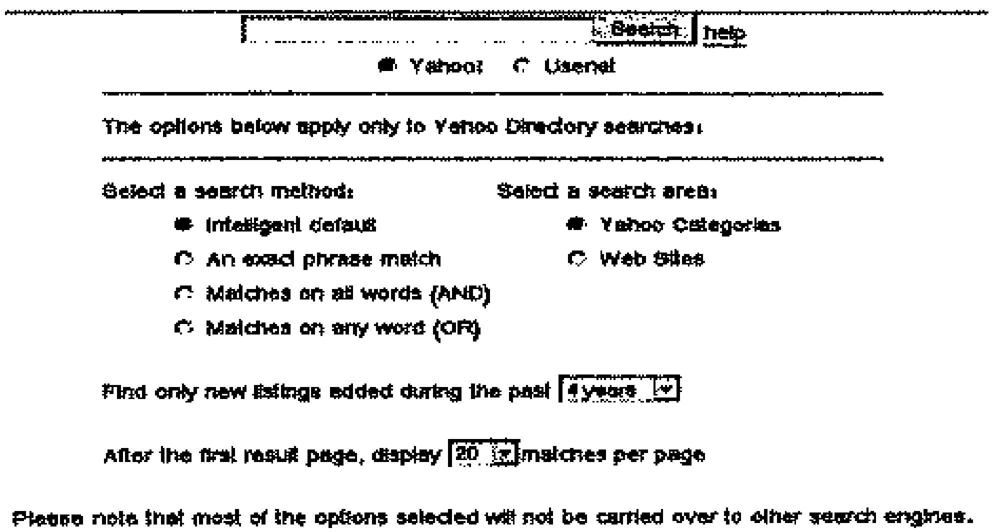
يُتيح Yahoo! للمستخدم أن يحدد مجموعة من الخيارات من أجل تعديل طريقة البحث والإظهار. يمكن القيام بذلك عن طريق الانتقال إلى صفحة الخيارات (Options) من خلال اتباع الارتباط التشعبي Advanced Search، ثم التعديل فيها بما يناسبه. يبين الشكل (6-4) أنواع الخيارات في صفحة Options.

الخيارات الموجودة هي:

- تحديد طريقة البحث، أو طريقة معاملة طلبات المستخدم، أو تحديد نوع العلاقة بين كلمات الطلب، وهي واحدة من العلاقات التالية.
  - علاقة "و AND"، أي أن الصفحات المطلوبة يجب أن تحتوي جميع كلمات الطلب.
  - علاقة "أو OR"، أي أن الصفحات المطلوبة تحتوي إحدى كلمات الطلب أو مجموعة منها.
  - التعامل مع الطلب كما لو كان جملة واحدة لا تتجزأ.

- الوضع الافتراضي "الذكي" (Intelligent Default)، حيث يبحث عن الصفحات التي تحوي أكبر عدد من الكلمات التي ضمن الطلب.

- ☐ تحديد مجال البحث، هل هو فئات Yahoo! أم مواقع Yahoo!.
- ☐ تحديد مكان البحث، هل هو ضمن Yahoo! أم ضمن Usenet (مجموعة المجموعات الإخبارية).



الشكل (6-4): صفحة Options التي تسمح بتعديل خيارات البحث باستخدام Yahoo!

تقييم البحث بالمحتوى

1. البحث البسيط

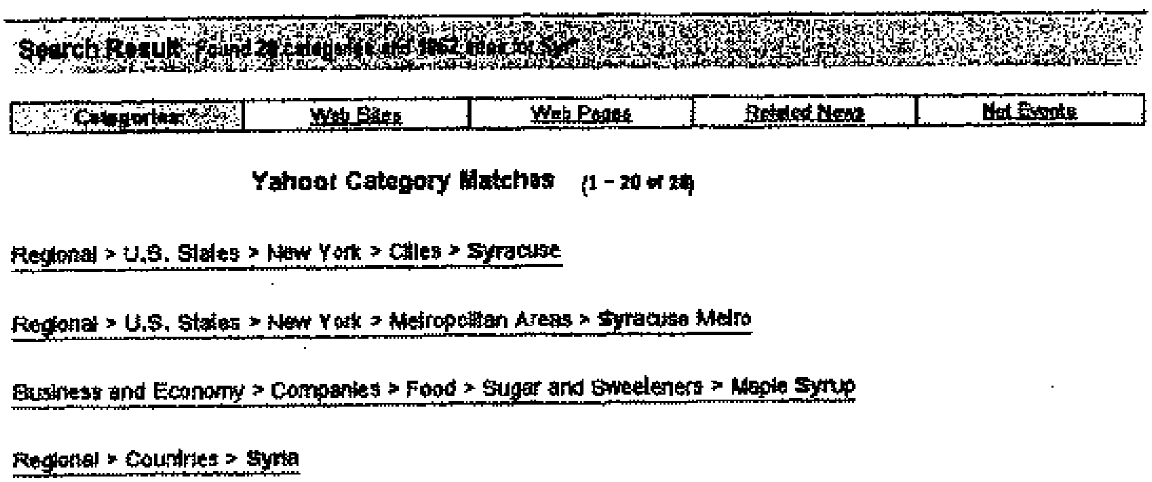
يسمح Yahoo! بعمليات البحث البسيط، ومن أجل فحص إمكانات هذا النوع من البحث، قمنا بالتجارب التالية:

- التحسس للأحرف الكبيرة والصغيرة: بإجراء البحث عن "Tourism" و"tourism" نجد أن عدد الصفحات التي يجدها النظام لا يختلف بين الحالتين، وهذا يدل على أنه لا يتحسس للأحرف الصغيرة والكبيرة.
- البحث عن الجمل: بإجراء البحث عن "tourism in Syria"، نجد أن جميع الصفحات التي وجدها النظام تتحدث عن السياحة في سورية، وتحتوي هذه الجملة. هو إذن يتيح إمكانية تجميع كلمات البحث في جمل، إذ يكفي وضع الجملة ضمن "" حتى يقوم بعملية بحث عن الصفحات التي تحوي الجملة بكاملها.
- الكلمات المفتاحية اللازمة (Required Keywords): يتيح Yahoo! للمستخدم تحديدَ كلمة يريد أن تكون محتواة ضمن جميع الصفحات المطابقة باستخدام إشارة "+" أمام الكلمة اللازمة. فمثلاً بالبحث عن tourism +Syria نجد الصفحات التي تتحدث عن tourism وتحتوي ضمنها بالضرورة Syria.
- الكلمات المفتاحية الممنوعة (Prohibited Keywords): يتيح Yahoo! إمكانية منع كلمة من أن تكون موجودة ضمن صفحة ما عن طريق استخدام "–". فمثلاً بعد إجراء البحث عن "tourism" والبحث عن "tourism -Syria"، نجد بفحص النتائج أن الصفحات التي تتحدث عن "Syria" قد اختفت في الطلب الثاني.
- يسمح Yahoo! باستخدام الرمز * (Wild Character) في آخر الكلمة فقط من أجل الدلالة على مجموعة غير محددة من الأحرف. فمثلاً بالبحث "Syr*"، نجد جميع الصفحات التي تحوي كلمات تبدأ بـ "Syr". يُظهر الشكل (6–5) النتائج التي حصلنا عليها (الشكل معالج ومعدّل من أجل عرض الجزء الهام منه فقط).
- يسمح Yahoo! بترتيب الأجوبة على طلب المستخدم وفق المعايير التالية:

يجد أولاً جميع الصفحات المطلوبة، ثم يرتب صفحات كل مجال على حدة وفق المنطق التالي:

- تُعطى الصفحات درجات حسب عدد الكلمات التي تحويها من بين الكلمات المستخدمة في الطلب، وكلما احتوت الصفحة كلمات أكثر، كلما كانت الدرجة أعلى.
- تُعطى الصفحات التي ترد إحدى كلمات البحث ضمن عنوانها أو ضمن الـ URL الخاص بها درجات أعلى.
- تُعطى الصفحات التي تنتمي إلى الفئات (Categories) الأكثر عمومية درجات أعلى.

إن طريقة الترتيب هذه بسيطة وغير كافية، فهي لا تقدم معلومات كثيرة للمستخدم الذي قد يضطر في كثير من الأحيان أن يتصفح عدداً كبيراً من الوثائق قبل أن يعثر على الوثيقة المطلوبة.



الشكل (6-5): استخدام * في كتابة الطلبات

2. البحث المتقدم

لا يسمح Yahoo! باستخدام أي نوع من أنواع البحث المتقدم، وهو أمر يؤثر على جودته كمحرك بحث.

3. الفلترة

يسمح Yahoo! بنوع وحيد من الفلترة: حسب التاريخ فقط، إذ يمكن أن نحدد الصفحات التي حُرِّرت منذ تاريخ معين وحتى الآن. يمكن تعديل هذا التاريخ من خلال صفحة الخيارات (Options) المذكورة آنفاً.

4. التصنيف

يؤمن Yahoo! تصنيفاً جيداً للمواضيع ضمن فئات، وهذا يساعد في تضييق مجالات البحث أكثر، وبالتالي يفيد في تسريع الوصول إلى المعلومات.

5. إظهار النتائج

تظهر نتائج Yahoo! على طلب المستخدم ضمن خمسة أقسام، كل قسم مستقل عن الآخر، ويمكن الانتقال بينها، كما نلاحظ في الشكل (6–6).



شكل (6–6): أقسام نتائج البحث في Yahoo!

وهكذا تُقسم نتائج البحث حسب أقسام قاعدة البيانات التي ذكرناها آنفاً، أي حسب المكان الذي يبحث فيه Yahoo! عن عناوين مطابقة لطلب المستخدم، وهي:

آ– فئات *Yahoo!*

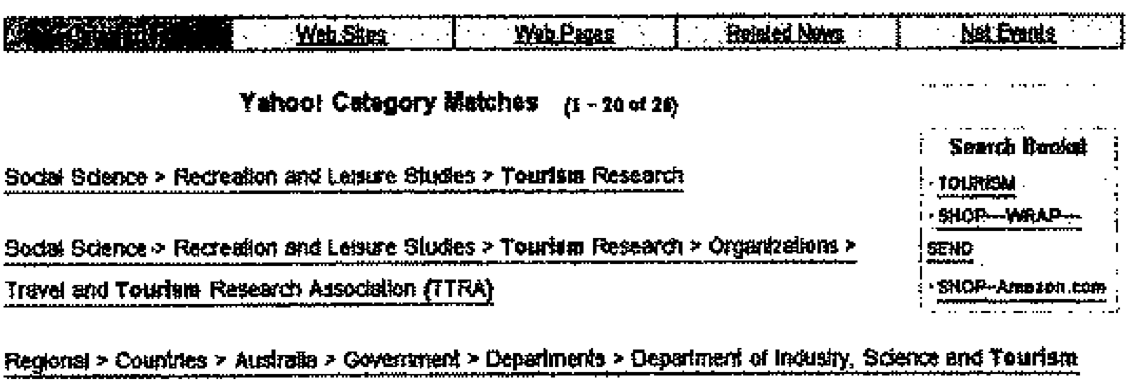
وهي مجموعة الفئات التي وجدها Yahoo! في قاعدة بياناته والمطابقة لطلب المستخدم. نلاحظ في بداية صفحة النتائج (الشكل 6–7)، أن Yahoo! يعطينا عدد الفئات المطابقة التي وجدها إضافةً إلى عدد المواقع التي وجدها. إن سبب وجود المعلومة الأخيرة هنا هو آلية بحث Yahoo!، فهو يبحث ضمن الفئات وضمن المواقع بنفسه قبل أن يطلب من Inktomi أن يتابع البحث.



الشكل (6–7): عدد نتائج البحث من الفئات والمواقع

يُظهر Yahoo! مثلاً نتائج البحث عن "Tourism" في الشكل (6–8) الذي نلاحظ فيه الأمور التالية:

- كلمة Categories مظللة دلالةً على أننا ضمن قسم نتائج Categories.
- يُحدد Yahoo! عدد النتائج التي يُظهرها على الصفحة الحالية (هنا 20 في كل صفحة)، والعدد الكلي (هنا 26 نتيجة).
- النتيجة هي ارتباط تشعبي إلى الفئة المطابقة لطلب المستخدم. نجد في الشكل أن الفئة الأولى المطابقة هي Tourism Research، وهي فئة تخصصية من الفئة Recreation and Leisure Studies، أما الفئة العليا والأعم فهي Social Science. ونلفت انتباه القارئ إلى أن باستطاعته الوصول إلى هذه الوثيقة من خلال التجول بين فئات Yahoo!، كما ذكرنا آنفاً.
- تظهر كلمات الطلب ضمن النتيجة بلون أسود (Bold).
- نجد على يمين الشكل مربعاً بلون مغاير بعنوان Search Books!. يقوم Yahoo! من تلقاء نفسه بالبحث عن مجموعة كتب مطابقة لطلب المستخدم، ويقدمها له إن أحب الاطلاع عليها.

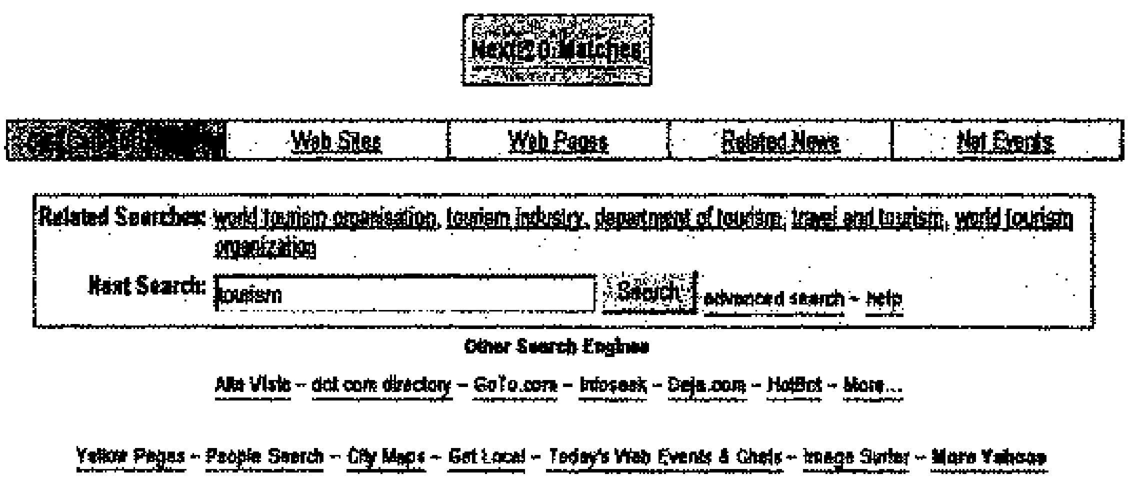


الشكل (6–8): نتائج البحث من الفئات

- في نهاية الصفحة، نجد ارتباطاً تشعبياً إلى باقي نتائج البحث بعنوان Next 20 Matches، ونجد من جديد إمكانية الانتقال إلى باقي أقسام نتائج البحث. يلي ذلك مجموعة من الارتباطات المتعلقة بطلب المستخدم (Related Searches)، ونجد في الشكل (6-9) مجموعة من هذه الارتباطات مثل world tourism organization (منظمة السياحة العالمية)، وغيرها...

- تلي ذلك إمكانية صياغة طلب جديد، مع نفس الإمكانات السابقة، فيما إذا وجد المستخدم أن الطلب الحالي غير كامل، أو أنه عام جداً، أو أن النتائج التي وجدها Yahoo! غير مقنعة..

- أخيراً، وفي أسفل الصفحة (الشكل 6-9)، يقدم Yahoo! إمكانية البحث عن موضوع هذا الطلب باستخدام محركات بحث أخرى، أسماؤها مدرجة كارتباطات تشعبية أو أدوات بحث أخرى ضمن Yahoo! (مثل Yellow Pages)، ويكفي اتباع الارتباط ليقوم Yahoo! بتحويل الطلب إلى محرك البحث الموافق، ولن يحتاج المستخدم إلى إعادة صياغة الطلب لأن Yahoo! يقوم بالمهمة. نلفت انتباه القارئ إلى أن عملية نقل الطلب لا تتضمن نقل خيارات البحث التي ضبطها المستخدم ضمن صفحة Options.



الشكل (6-9): إمكانية إجراء بحث جديد

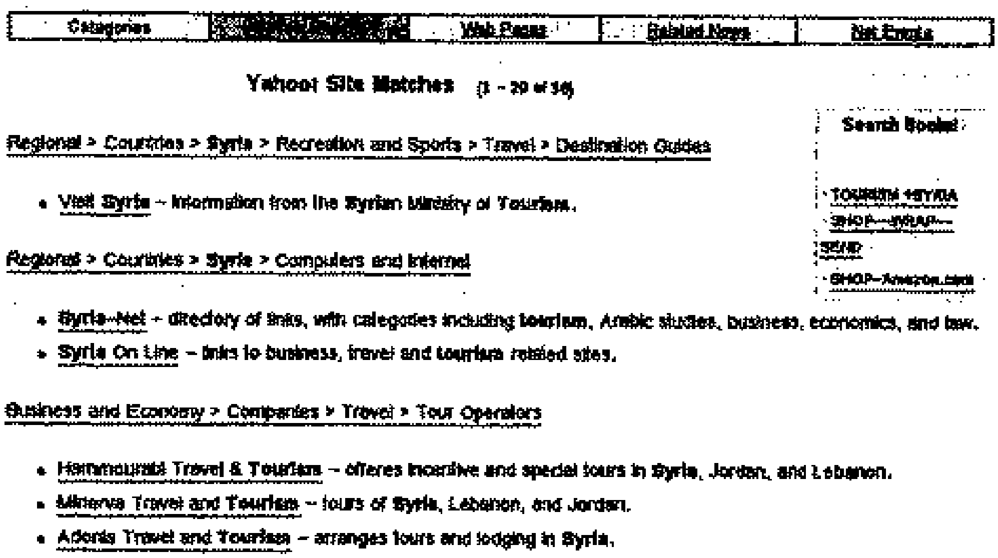
ب- *مواقع الوب*

وهي مجموعة المواقع التي وجدها Yahoo! ضمن قاعدة بياناته والمطابقة لطلب المستخدم. فمثلاً إذا بحثنا عن tourism + Syria، نجد أن عدد الفئات المطابقة 0، وعدد المواقع المطابقة 30 (الشكل 6-10). أمّا الجديد هنا في صفحة النتائج الحالية مقارنةً بسابقتها (الشكل 6-11) فهو ما يلي:



الشكل (6-10): عدد نتائج البحث من الفئات والمواقع

تظهر النتيجة كارتباط تشعبي عريض يدل على الفئة الأكثر قرباً من طلب المستخدم (وفي حالتنا لا توجد فئة مطابقة لطلب المستخدم)، وأسفل هذا الارتباط التشعبي توجد ارتباطات تشعبية أخرى إلى مجموعة المواقع المطابقة لطلب المستخدم. ففي الشكل (6-11)، نلاحظ أن Syria هي إحدى فئات Yahoo!، وهي فئة فرعية من Countries، أما الفئة الأكثر قرباً إلى طلب المستخدم فهي Destination Guides وهي فئة تخصصية من Syria. نجد، أسفل هذا الارتباط، الموقعَ المطابق لطلب المستخدم، والذي يمكن الوصول إليه عن طريق الارتباط التشعبي Visit Syria، وبجانبه تعليقاً صغيراً عن هذا الموقع يدل على أن الموقع يحتوي معلومات وضعتها وزارة السياحة السورية عن السياحة في سورية.



Categories | | Web Pages | Related News | Net Events

Yahoo! Site Matches (1 - 20 of 30)

Search Books:

TOURISM +SYRIA
SHOP--WRAP--
SEND
SHOP--Amazon.com

Regional > Countries > Syria > Recreation and Sports > Travel > Destination Guides

- Visit Syria - Information from the Syrian Ministry of Tourism.

Regional > Countries > Syria > Computers and Internet

- Syria-Net - directory of links, with categories including tourism, Arabic studies, business, economics, and law.
- Syria On Line - links to business, travel and tourism related sites.

Business and Economy > Companies > Travel > Tour Operators

- Hammoud Travel & Tourism - offeres incentive and special tours in Syria, Jordan, and Lebanon.
- Minerva Travel and Tourism - tours of Syria, Lebanon, and Jordan.
- Adonis Travel and Tourism - arranges tours and lodging in Syria.

الشكل (6-11): نتائج البحث من المواقع

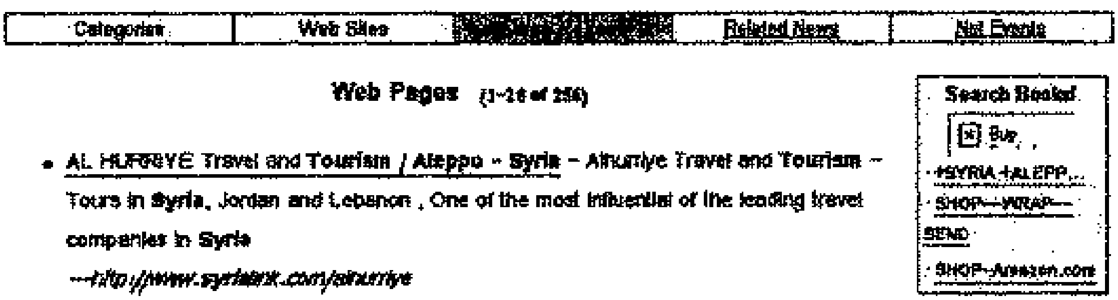
ج- صفحات الوب

وهي مجموعة الصفحات المطابقة لطلب المستخدم والتي يعيدها Inktomi. نجد في بداية صفحة النتائج الشكل (6–12). حيث أن الطلب المستخدم هنا هو syria aleppo tourism مع تحديد طريقة البحث "Match on all words (AND)" في صفحة Options. أعاد Inktomi 256 صفحة مطابقة لطلبنا السابق، وقام بعرضها كما في الشكل (6–13)، وقد قمنا بعرض النتيجة الأولى فقط من النتائج السابقة.



الشكل (6–12): عدد نتائج البحث من صفحات الوب

الاختلاف الوحيد في طريقة عرض النتيجة، هو أن الارتباط التشعبي على اسم الصفحة يقود مباشرةً إلى الصفحة الموافقة لطلب المستخدم، ثم يذكر Yahoo! تعليقاً على هذه الصفحة، ويضع عنوان الصفحة URL بشكل صريح.



الشكل (6–13): نتائج البحث من صفحات الوب

د- الأخبار الموافقة

وهي مجموعة من الأخبار المتعلقة بطلب المستخدم يجمّعها Yahoo! من قاعدة بياناته. في بداية صفحة النتائج (الشكل 6–14) نجد، كالعادة، عدد مقالات الأخبار التي وجدها Yahoo!. فمثلاً من أجل الطلب tourism، وجدنا مجموعة من المقالات عددها 88 مقالاً تتحدث عن السياحة tourism.

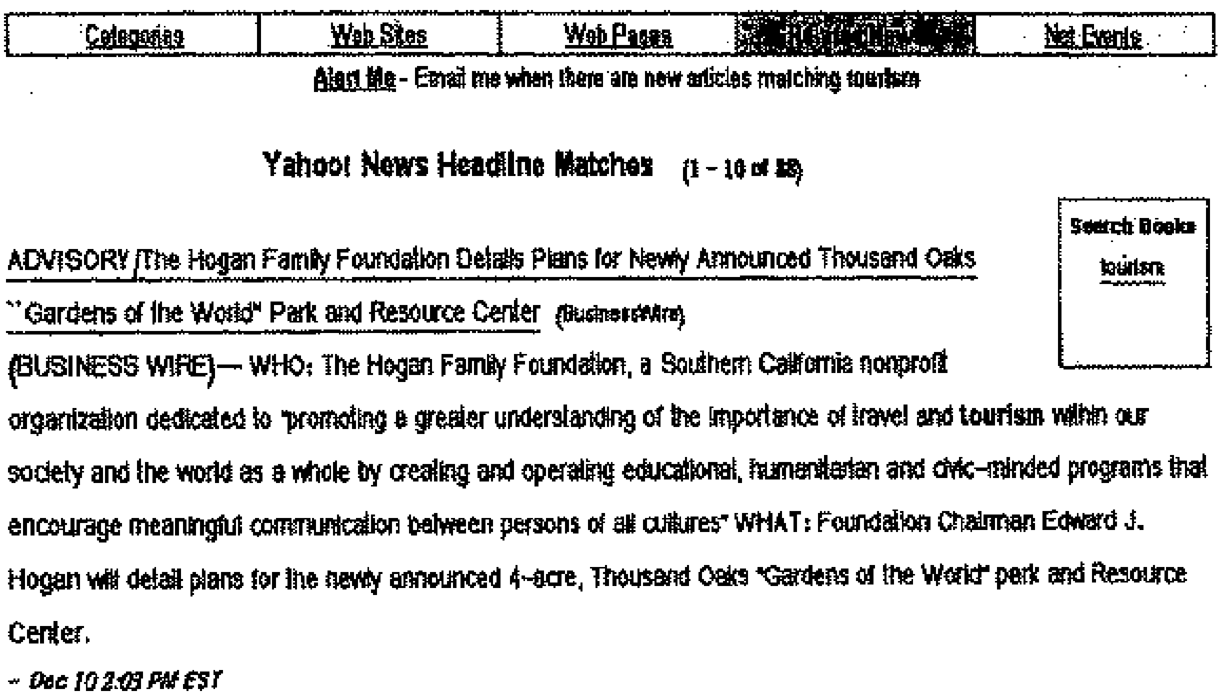


الشكل (6-14): عدد نتائج البحث من المقالات الإخبارية

أما النتائج فكان لها الشكل (6-15).

تظهر النتيجة على شكل ارتباط تشعبي على عنوان المقال، ثم مصدر المقال (وهو هنا BusinessWire)، ثم جزء من المقال يحدد مضمونه. نلاحظ أيضاً أن كلمات الطلب التي ترد ضمن الجزء المعروض من المقال تظهر بلون أسود.

نلاحظ في أسفل الصفحة أيضاً وجود مجموعة من الارتباطات التشعبية تسمح بالانتقال إلى مواقع تزوّد بالأخبار فيما إذا أردنا توسيع البحث عن مقال معين. ونرى في الشكل (6-16) وجود ارتباط تشعبي إلى CNN News وWashington Post.


Categories | Web Sites | Web Pages | Net Events

Alert Me - Email me when there are new articles matching tourism

Yahoo! News Headline Matches (1 - 10 of 88)

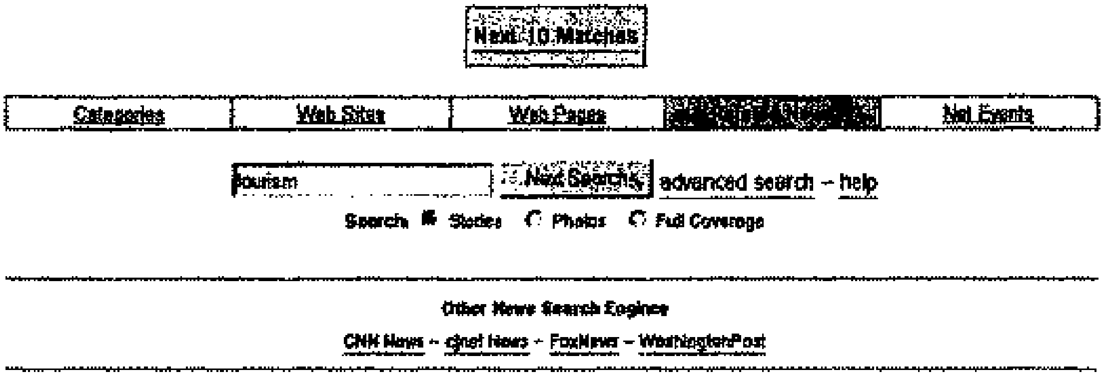
Search Books
tourism

ADVISORY/The Hogan Family Foundation Details Plans for Newly Announced Thousand Oaks "Gardens of the World" Park and Resource Center (BusinessWire)

(BUSINESS WIRE)— WHO: The Hogan Family Foundation, a Southern California nonprofit organization dedicated to "promoting a greater understanding of the importance of travel and **tourism** within our society and the world as a whole by creating and operating educational, humanitarian and civic-minded programs that encourage meaningful communication between persons of all cultures" WHAT: Foundation Chairman Edward J. Hogan will detail plans for the newly announced 4-acre, Thousand Oaks "Gardens of the World" park and Resource Center.

- Dec 10 2:03 PM EST

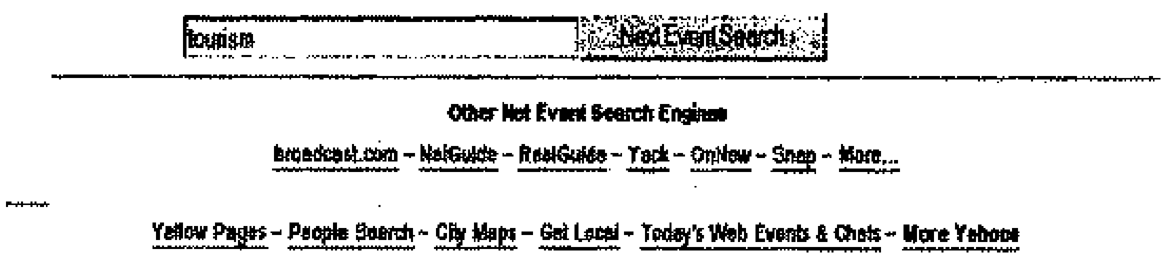
الشكل (6-15): نتائج البحث من المقالات الإخبارية



الشكل (6–16): إمكانات البحث في محركات الأخبار الأخرى

هـ – أحداث الشبكة

تعطي أحداث الشبكة مجموعة الأحداث الموجودة ضمن قاعدة البيانات، والتي تتلاءم مع طلب المستخدم. المميز في النتائج هنا (الشكل 6–17) هو استخدام محرك بحث خاص بـ Net Events، كما نجد مجموعة من الارتباطات التشعبية إلى مجموعة من محركات البحث الخاصة بـ Net Events مثل BroadCast وNetGuide وغيرها...



الشكل (6–17): استخدام محرك أحداث الشبكة

6. التغذية الراجعة

لا يتيح Yahoo! أية تغذية راجعة.

7. السرعة

يتمتع Yahoo! بسرعة جيدة، فلا نعاني من الانتظار حتى نحصل على النتائج.

8. البحث في العناوين والـ URL

يستطيع Yahoo! تحديد مجال البحث بعناوين الصفحات (Page Titles) فقط أو بـ URL، كما يبين المثالان التاليان: u:Syria, t:Syria. فاستخدام :t في المثال الأول يعني الحصول على الصفحات التي يحتوي عنوانها كلمة Syria (الشكل 6–18)، أمّا استخدام :u في المثال الثاني فيعني العثور على الصفحات التي يحتوي الـ URL الخاص بها كلمة Syria ضمنه (الشكل 6–19). كما يسمح أيضاً بدمج العمليات المختلفة كما في الطلب: t:tourism +Syria الذي يبحث عن الصفحات التي تحتوي كلمتي tourism وSyria ضمن عنوانها.

ج. البحث حسب الموصفات

يتيح Yahoo! العديد من الخدمات التي يمكن البحث فيها بطلبات من نمط تلك التي تُستخدم في قواعد بيانات مثل: البحث عن الأشخاص (Person Search)، والصفحات الصفراء (Yellow Pages) وسنتكلم عنها لاحقاً في هذا الفصل. المبدأ الأساسي فيها أن Yahoo! يعرض للمستخدمين واجهة تخاطبية (Form) فيها العديد من الحقول التي تمثل موصفات الخدمة (مثل الاسم الثاني لشخص إذا كنا نستخدم خدمة البحث عن الأشخاص). يستطيع المستخدم أن يكتب فيها القيم أو الكلمات التي يريد البحث عنها، عندئذ يبحث Yahoo! في قاعدة بياناته عن القيم الموافقة ويعرض النتائج في جداول.

Yahoo! Site Matches (1 – 23 of 39)

Regional > Countries > Syria > Recreation and Sports > Travel > Destination Guides

- Syria Guide – restaurants, hotels, entertainment, transportation, shopping, and other travel resources.
  http://www.syria-guide.com/
- Visit Syria – information from the Syrian Ministry of Tourism.
  http://www.visit-syria.com/
- Ministry Of Tourism – includes information for visitors.
  http://www.syriatourism.org

الشكل (6–18): البحث من كلمة Syria في الـ URL (u:Syria)

Yahoo! Category Matches (1 – 11 of 11)

Regional > Countries > Syria

Recreation > Travel > By Region > Countries > Syria

Business and Economy > Companies > Travel > Lodging > By Region > Countries > Syria

Recreation > Travel > Travelogues > By Region > Countries > Syria

Government > Politics > By Region > Countries > Syria

Business and Economy > Organizations > By Region > Countries > Syria

News and Media > Newspapers > By Region > Countries > Syria

News and Media > By Region > Countries > Syria

الشكل (6–19): البحث من كلمة Syria في العنوان (t:Syria)

6–1–3 خدمات المعلومات الأخرى

يتيح yahoo! العديد من الخدمات والأدوات، ومعظمها مجاني، وهي تساعد المستخدمين على الحصول على معلومات ذات طابع تخصصي، أو الاتصال بمستخدمين آخرين، أو الحصول على بريد إلكتروني ... نذكر فيما يلي بعضاً من أهم هذه الخدمات:

I. World Yahoos : أداة تسمح بالتجول عبر العالم من خلال ارتباطات تشعبية إلى مجموعة كبيرة من الدول. كما توجد ارتباطات إلى مواقع مدن ضمن الولايات المتحدة.

2. Mail: أداة مجانية تتيح للمستخدم إنشاء صندوق بريد خاص به على مخدّم Yahoo!، مما يتيح له الاتصال مع العالم باستخدام البريد الإلكتروني E-mail.

3. My Yahoo: أداة تتيح للمستخدم تخصيص Yahoo! وذلك بتجميع الأجزاء التي تهمه منه ضمن صفحة واحدة، مثل الرياضة والأخبار والطقس والأفلام والموسيقى وغيرها...

4. Health: تقدم معلومات صحية للمستخدمين، وتسمح أيضاً بطرح أسئلة على أخصائيين في الصحة، والحصول على الأجوبة بسرعة. يمكن أيضاً الانضمام إلى نوادٍ خاصة بالصحة لمناقشة أمور صحية مع الآخرين.

5. Shopping: تسمح بشراء الكثير من الحاجيات عن طريق الشبكة، فيمكن البحث عن أي منتج، والتجول بين المنتجات الجديدة، والتعرّف على متاجر جديدة.

6. Classifieds: وهي أداة مجانية تسمح بالبحث عن عمل أو أغراض للبيع (منازل، سيارات، ...)، كما تؤمن خدمة صندوق البريد mailbox المجانية. هذه الأداة مضمونةٌ (سرية صندوق البريد)، وسهلة الاستخدام، كما يمكن للمستخدم إضافة معلومات خاصة به إلى القائمة الموجودة.

7. News: وهي تعرض مجموعة من أحدث الأخبار في العالم. تُقسم هذه الأخبار (حسب أصنافها) إلى مجموعة من الفئات عددها 13، فنجد مثلاً الأخبار المالية والرياضية وغيرها.

8. Sports: تسمح بالاطلاع على مختلف الألعاب الرياضية والمباريات والنتائج والإحصاءات، إضافةً لإمكانية جمع معلومات عن فريق معين، وغير ذلك مما يتعلق بالرياضة وأخبارها.

9. Yellow Pages: وهي صفحات موزعة ضمن فئات، تحتوي على معلومات عن الأعمال الموجودة ضمن مناطق مختلفة من العالم. يمكن التجول بين هذه الصفحات، فمثلاً يمكن أن نتجول بين مجموعة الصفحات التي تتحدث عن الأعمال ضمن مدينة معينة، فإن لم نجد ما نطلبه، نستطيع الانتقال إلى مدينة مجاورة والبحث عما نطلبه ضمنها. يمكن أيضاً صياغة طلبات بحث عن مادة أو

نشاط اقتصادي في مدينة معينة، بأن نحدد رمز المدينة، والمادة التي نطلبها. يمكن أيضاً البحث ضمن مدن قريبة جغرافياً من مدينة ما.

10. Maps and Driving Directions: وهي أداة تُظهر خرائط للمناطق، ويمكن باستخدامها العثور على الطريق الذي يصل بين نقطتين من مدينة مثلاً. الخرائط المتاحة هي خرائط ضمن الولايات المتحدة فقط.

11. People Search: وهي أداة تسمح للمستثمر بالبحث عن أشخاص والحصول على عنوانهم البريدي والإلكتروني، ورقم هاتفهم، وغير ذلك من المعلومات.

12. Small Business: أداة مفيدة للمقاولين ولمن يديرون أعمالاً صغيرة. تسمح بمختلف أنواع الاتصالات التي يتيحها Yahoo! عبر الشبكة، كما تتيح مجموعة أخرى من الأدوات والخدمات مثل إعطاء أسعار المواد ومعلومات عن السوق مع إمكانية تقديم فرص عمل ضمن Classifieds وغيرها...

13. Lodging: وهي أداة تزود السائح بمعلومات عن الفنادق أو منازل الإيجار التي يرغب بالنزول فيها للاستجمام. تشتمل هذه المعلومات على الإيجار وأنواع الطعام في مختلف الوجبات، وأرقام الهواتف والفاكس، و غيرها...

14. Travels: تؤمن جميع المعلومات اللازمة للسفر، من معلومات عن الرحلات، إلى تأمين الحجوزات، ومعلومات عن البلد المقصود، ومعلومات عن أسعار تحويل العملات، إضافةً إلى دليل سياحي، والمزايا المتاحة.

15. Parks: تعطي معلومات عن منتزهات وحدائق في جميع أنحاء العالم.

16. Site: يسمح yahoo! للمستخدم بإنشاء موقع وب خاص به على الشبكة، وبأجرة 30 دولاراً شهرياً، ويعطي لهذا الموقع العنوان WWW.*name*.com. ينشئ Yahoo! هذا الموقع على مخدّمه الخاص، ويقدم للمستخدم واجهات عديدة مع مجموعة من التقانات المختلفة كي يختار منها ما يناسبه.

6–1–4 إضافة موقع إلى محرك !Yahoo

يمكن للمستخدم أن يُعرِّف موقعه الخاص ضمن فئات !Yahoo على النحو التالي:

يختار المستخدم الفئة التي ينتمي موقعه إليها من الصفحة الرئيسية، ثم يختار الفئة الفرعية من تلك الفئة، ويستمر في اختيار الفئات الفرعية حتى يصل إلى الفئة التي تناسب موقعه فعلاً. يقوم عندئذ باختيار الارتباط التشعبي "Suggest a Site" حيث تظهر له مجموعة حقول يقوم بملئها بمعلومات عن الموقع، مثل URL ووصف الموقع وطريقة الاتصال والكلمات المفتاحية وغيرها. بعد ذلك يقوم أشخاص مختصون بزيارة هذا الموقع وإضافته إلى "مجلد !Yahoo" (Yahoo! Directory).

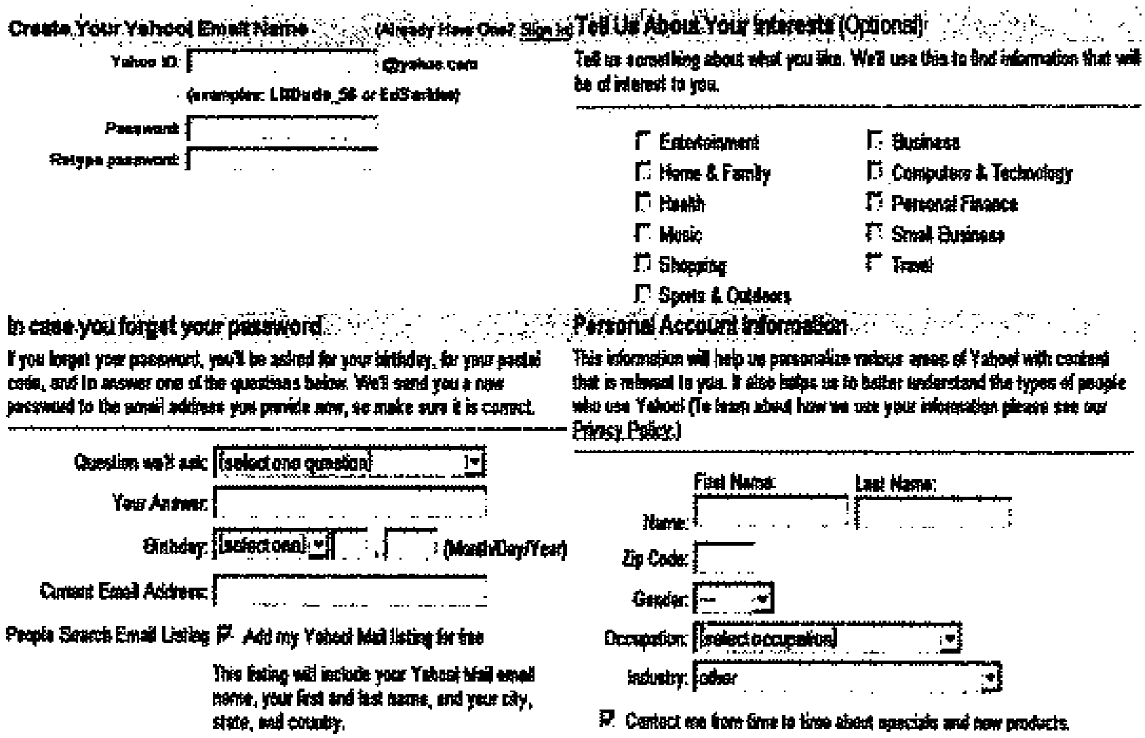
لما كان العمل يدوياً بالكامل، فقد ينتظر المستخدم يومين حتى يجد أن موقعه قد عُرِّف ضمن !Yahoo ويصبح بمقدوره الوصول إليه باستخدام عمليات بحث مناسبة.

6–1–5 الانضمام إلى مجتمع !Yahoo

يستطيع !Yahoo أن يتعامل مع المستخدمين كما لو كان لكل منهم حساب خاص فيه، فيعطي المستخدم إمكانية الحصول على رمز خاص به، وكلمة سر. يستخدم !Yahoo الرمز وكلمة السر في جميع الأدوات والخدمات التي يطلبها المستخدم، فيستطيع مثلاً استخدام Mail مع نفس الرمز ونفس كلمة السر. تجري عملية التسجيل في !Yahoo كما يلي:

يتبع المستخدم الارتباط التشعبي Sign Up Now، فتظهر أمامه صفحة فيها مجموعة من حقول الحوار. يجب عليه ملء هذه الحقول التي تطلب معلومات عن الرمز الذي يريده وكلمة السر، ثم معلومات شخصية مثل الاسم الكامل والرمز المختصر (Zip Code) والجنس والمهنة وميدان العمل، ومعلومات أخرى عن اهتماماته. يوضّح الشكل (6–20) هذه الحقول، مع ملاحظة أن الواجهات ليست مرتبة بهذا الشكل، بل هي مرتبة بعضها تحت بعض.

يوجد بين المعلومات التي يطلبها Yahoo! مجموعة حقول بعنوان "In case you forget your password". تُستخدم هذه المجموعة إذا أضاع المستخدم كلمة السر التي يستخدمها، إذ يمكن عندها الطلب من Yahoo! استرجاعها. يتحقق Yahoo! من هوية الشخص الذي يطلب كلمة السر بأن يسأله السؤال الذي (حدّده المستخدم بنفسه) ضمن الحقل الأول، كما يطلب منه تحديد تاريخ ميلاده، فإن كانت الأجوبة مطابقة لما لديه، يعطي Yahoo! المستخدم كلمة السر. إذا كان لدى المستخدم E-mail غير ذلك الذي أعطاه إياه Yahoo!، يرسل النظام كلمة السر إلى ذلك العنوان.



الشكل (6–20): واجهات طلب معلومات من أجل تسجيل مستخدم جديد في yahoo!

6-2 محرك البحث HotBot

WWW.HotBot.com

من أحدث المحركات على الشبكة ويُنتظر أن يكون من أفضلها. تكمن قوته الأساسية في البحث ضمن مجموعات الأخبار.

يُفهرس HotBot مجموعة كبيرة من الصفحات تقدّر بـ 110 مليون صفحة، كما أن لديه مجلداً خاصاً مقسماً إلى فئات (شبيه بمجلد Yahoo!). يُبنى هذا المجلد يدوياً، ويشتق معلوماته من المجلد المفتوح (Open Directory)، وهو أضخم مجلد تصنيف يدوي للمواقع على الشبكة في العالم، ويشترك فيه محركا البحث HotBot وLycos.

من المفيد هنا الإشارة إلى بعض الخواص المفيدة لمجلد HotBot، مثل:

- عدد الارتباطات التشعبية المكثفة ضمن فئة معينة.
- ارتباطات إلى فئات متعلقة (Related Categories) بالفئة الحالية (Current Category) ضمن المجلد. تفيد في توسيع أو تضييق بحث المستخدم بالتجول في فئات معينة.
- مجموعة من المواقع (Site recommendations) حُدِّدت من قبل خبراء الفئة الحالية ضمن المجلد على أنها مواقع ذات أهمية خاصة لمن يتجول في هذه الفئة.
- مجموعة من الارتباطات التشعبية إلى محرّري الفئة (Category editors) الحالية، أي أولئك الذين يهتمون بتنظيم هذه الفئة.
- الارتباط التشعبي Submit a Site الموجود في جميع فئات المجلد عدا الفئات ذات الصفة العمومية. يسمح هذا الارتباط للمستخدم بتسجيل موقع ضمن المجلد في فئة معينة (الفئة التي

اتبع منها الارتباط التشعبي). طبعاً لن يُضاف هذا الموقع مباشرةً، بل يُدرس من قبل محرّري الفئة، ثم يضاف إلى الفئة المناسبة. يُكتب بجانب المواقع المضافة حديثاً كلمة New باللون الأحمر.

- الارتباط التشعبي Become an Editor الذي يسمح للمستخدم بأن يُصبح محرّر فئة معينة بعد أن يملأ استمارة خاصة.

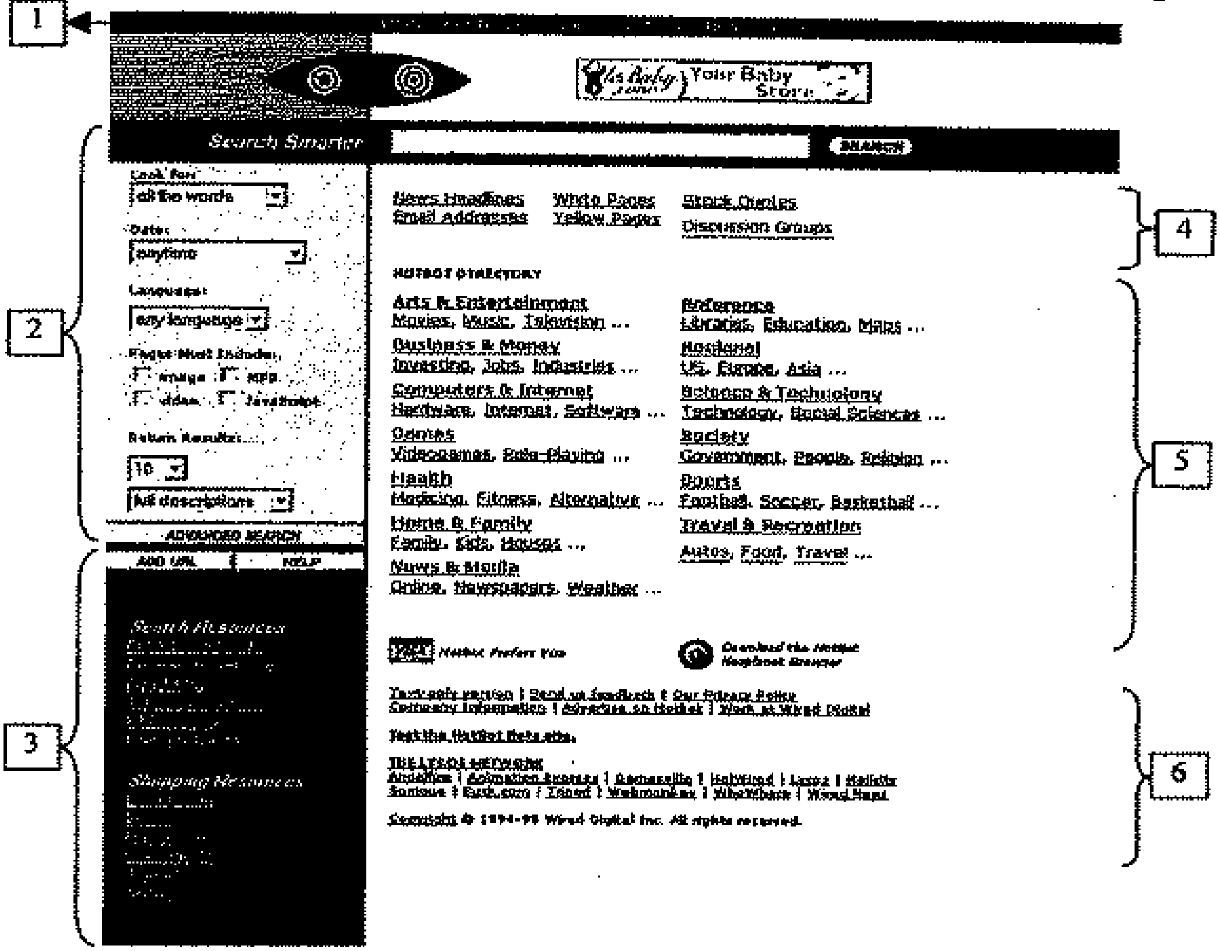
6-2-1 واجهة HotBot

يُظهر الشكل (6-21) صفحة HotBot على الويب والتي نصل إليها باستخدام العنوان http://www.hotbot.com. تحتوي هذه الصفحة جميع إمكانات البحث التي يقدمها HotBot والعديد من خدمات المعلومات الأخرى. نجد فيما يلي الإمكانات المختلفة لـ HotBot:

- الجزء [1]: مجموعة من الارتباطات التشعبية إلى "جزء من شبكة Lycos".
- الجزء [2]: واجهة البحث في HotBot.
- الجزء [3]: مجموعة موارد HotBot.
- الجزء [4]: مجموعة ارتباطات تشعبية إلى مجموعة من أدوات البحث التي يتيحها HotBot، وسنتحدث عنها لاحقاً.
- الجزء [5]: مجلد HotBot، وهو يسمح بالبحث بالتصفح.
- الجزء [6]: مجموعة من الارتباطات التشعبية المختلفة إلى مجموعة من المواقع والأدوات. ونلفت الانتباه إلى أول ارتباط تشعبي Text-only Version، فهو يؤمن نسخة من HotBot لا تحتوي سوى نصوص فقط، مما يمكّن من استخدام HotBot في مستعرض نصوص (أي لا يدعم الوسائط المتعددة الأخرى).

6-2-2 البحث عن المعلومات في HotBot

يتيح HotBot الأنواع الثلاثة من البحث عن المعلومات: التصفح، البحث عن المحتوى، البحث عن الموصفات.



الشكل (6-21): صفحة HotBot

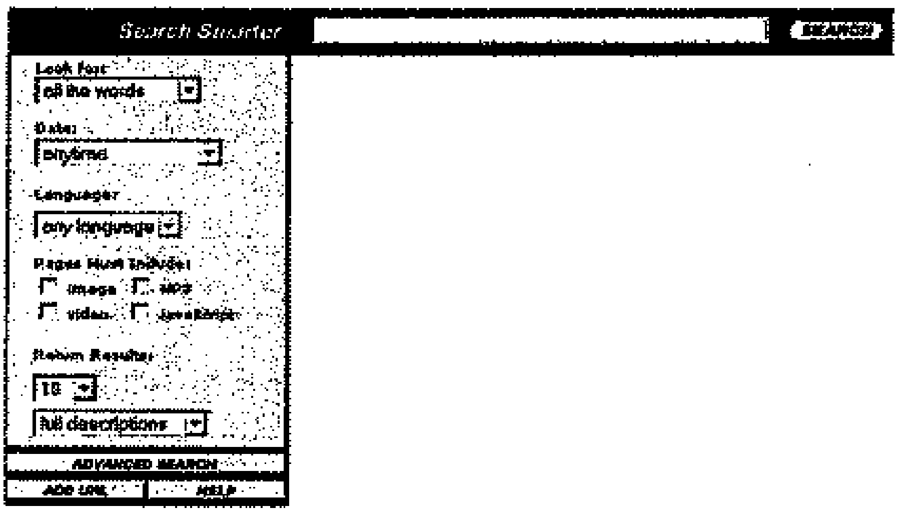
آ. التصفح

لا يختلف البحث بالتصفح في HotBot عنه في Yahoo!، ففي الجزء 5، نجد مجموعة فئات مجلد HotBot ذات المستوى الأعلى، ومنها يمكن الوصول إلى الفئات الفرعية الأكثر تخصصاً.

ب. البحث عن المحتوى

يتيح HotBot إمكانية صياغة طلبات بنفس الأسلوب المتبع في أنظمة استرجاع المعلومات، فيمكن صياغة الطلب ضمن الحقل الحواري الموجود في أعلى الجزء 3.

يمكن القيام بعمليات بحث بسيط باستخدام الصفحة الأساسية التي تؤمن مجموعة من إمكانات البحث المحدودة، كما يمكن استخدام عمليات بحث أكثر تعقيداً وتخصصاً من خلال الزر Advanced Search الذي نراه أسفل الجزء 2، وسنشرحه لاحقاً في هذا الفصل.



الشكل (6–22): الجزء 2 الذي يسمح بالبحث ضمن HotBot

تظهر في الشكل (6–22) مجموعة من الإمكانات التي تفيد في تحسين نتائج البحث، وهي:

- المنزلقة الأولى بعنوان Look for: تسمح بتحديد مجموعة من الخيارات حول طريقة معالجة الطلب وكلماته. هذه الخيارات هي:
  - Any of the words: اعتبار وجود OR بين جميع الكلمات.
  - All of the words: اعتبار وجود AND بين جميع الكلمات.

- Exact phrase: اعتبار الطلب كجملة واحدة.
- Page title: البحث ضمن عناوين الصفحات فقط.
- The person: اعتبار الطلب اسم شخص، وبالتالي البحث عن هذا الشخص.
- Links to this URL: يعتبر الطلب هنا هو URL، ويجري البحث عن جميع الارتباطات التشعبية إلى هذا العنوان.
- Boolean phrase: يعتبر الطلب هنا جملة منطقية، أي تحوي AND, OR, NOT ضمنها.

- المنزلقة الثانية بعنوان Date: تسمح بفلترة النتائج حسب التاريخ. فمثلاً تحديد الخيار "In the last 2 years"، يؤدي إلى إعادة جميع الصفحات التي تاريخ نشرها أصغر بسنتين من التاريخ الحالي.
- المنزلقة الثالثة بعنوان Language: تسمح بفلترة على لغة الصفحات النتائج. فمثلاً تحديد الخيار English يؤدي إلى إعادة جميع الصفحات المكتوبة باللغة الإنكليزية حصراً والمطابقة للطلب.
- مجموعة من الخيارات حول محتوى الصفحات النتائج، فيمكن تحديد إن كانت الصفحات الناتجة ستحوي صوراً أم لا، وغيرها من الخيارات الظاهرة على الشكل.
- المنزلقة الرابعة بعنوان Return Results: تحدد عدد النتائج التي تريد من HotBot أن يُظهرها في كل صفحة نتائج.

- المنزلقة الخامسة والأخيرة تؤثر على شكل عرض النتائج في صفحة النتائج، أي هل تريد إظهار URL الصفحات النتائج فقط، أم معه شرح[1] بسيط، أم شرح كامل.
- زر الانتقال إلى البحث المتقدم، وزر لتسجيل موقع، وآخر من أجل المساعدة.

من الجدير بالذكر أن HotBot يستخدم أيضاً محرك البحث الآلي Inktomi في عمليات البحث التي يقوم بها.

تقييم البحث بالمحتوى

1. البحث البسيط

يسمح HotBot بعمليات البحث البسيط، ومن أجل فحص إمكانات هذا النوع من البحث، قمنا بالتجارب التالية:

- التحسس للأحرف الكبيرة والصغيرة: البحث عن tourism و Tourism. تجد اختلافاً في عدد صفحات النتائج في كل مرة، مما يعني أنه يتحسس للكلمات الكبيرة والصغيرة. يتحسس HotBot للأحرف الصغيرة والكبيرة على النحو التالي: إذا وُجد حرف كبير ضمن الكلمة، فإن البحث يصبح حساساً للأحرف الصغيرة والكبيرة، أما إذا كانت جميع الأحرف صغيرة فإن البحث يصبح غير حساس للأحرف الكبيرة والصغيرة، أي أن البحث عن tourism يقبل كلمات مثل TOURISM أو Tourism، أما البحث عن Tourism فلا يقبل tourism أو TOURISM أو TouRism، بل يقبل Tourism فقط.
- البحث عن الجمل: البحث عن "tourism in Syria". نجد أن جميع الصفحات التي وجدها النظام تتحدث عن السياحة في سورية، وتحتوي هذه الجملة. إذن يتيح HotBot إمكانية

---

[1] يمكن للمستخدم أن يضيف هذا الشرح بنفسه باستخدام مؤثرات *HTML*، وسنرى هذا عند الحديث عن إضافة موقع إلى *HotBot*.

تجميع كلمات البحث في جمل، إذ يكفي وضع الجملة ضمن "" حتى يقوم بعملية بحث عن الصفحات التي تحتوي الجملة بكاملها. كما يمكن أن نبحث عن جملة بطريقة أخرى، وذلك بأن نحدد من المنزلقة الأولى (عنوانها look for) ضمن واجهة HotBot الرئيسية الخيار Exact Phrase، ثم نكتب الجملة دون استخدام "".

- الكلمات المفتاحية اللازمة: يسمح HotBot للمستخدم بأن يحدد كلمة يريدها أن تكون محتواة ضمن جميع الصفحات المطابقة. فمثلاً يؤدي الطلب: tourism +Syria إلى استحضار الصفحات التي تتحدث عن tourism وتحتوي ضمنها بالضرورة كلمة Syria.

- الكلمات المفتاحية الممنوعة: يتيح HotBot إمكانية منع كلمة من أن تكون موجودة ضمن صفحة ما عن طريق استخدام "-". فمثلاً: بالبحث عن "tourism" والبحث عن "tourism -Syria"، نجد بفحص النتائج أن الصفحات التي تتحدث عن "Syria" قد اختفت من الطلب الثاني.

- يسمح HotBot باستخدام الرمز * (Wild Card) ضمن كلمة من أجل مطابقة 0 حرف أو أكثر كما في الطلب: "*Syr" أو "yria*"، حيث يبحث عن جميع الصفحات التي تحتوي كلمات تبدأ بـ "Syr" أو تنتهي بـ "yria". كما يمكن البحث باستخدام ? من أجل مطابقة حرف واحد من الكلمة، مثلاً "s?n" يعيد "son" و"sun" و... يمكن استخدام الرمزين (* ,?) في أي مكان ضمن الكلمة، كما يمكن استخدامهما معاً ضمن نفس الكلمة، لكن يجب أن تتألف كلمة البحث من ثلاثة حروف على الأقل بما فيها الرمز (* ,?).

- يرتب HotBot النتائج حسب درجة مطابقتها، ويعطي درجة المطابقة كنسبة مئوية (لم نتمكن من مشاهدتها على صفحة النتائج). يقوم HotBot بحساب درجة المطابقة حسب مجموعة من العوامل أهمها:

  * ورود كلمات الطلب ضمن العناوين، فالصفحات التي تحتوي كلمات الطلب ضمن عناوينها تُعطى درجة مطابقة أعلى من تلك التي تحتوي الكلمات ضمن نصها.

- ورود كلمات الطلب ضمن مجموعة الكلمات المفتاحية التي كتبها محرِّر الصفحة في مؤشرات HTML الخاصة (Meta Tags) يعطي درجة مطابقة أعلى.
- تردد كلمات الطلب في النص، فالتردد العالي للكلمة يُعطي درجة مطابقة أعلى ؛ طبعاً يُراعى أن تكون الكلمات العامة مثل The أقل تأثيراً من الكلمات الخاصة مثل Alfred.
- طول الوثيقة.
- يقوم HotBot بتخفيض درجة مطابقة الصفحات التي تأتي من مواقع تسبب الإزعاج[3].

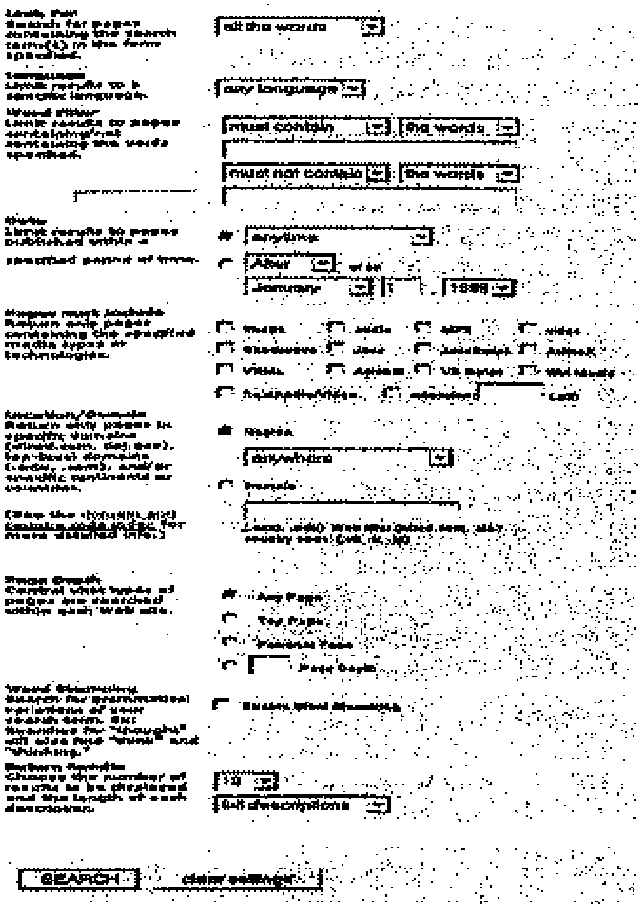
ونذكر أن HotBot يُحدد العدد 1000 كحد أقصى لعدد نتائج البحث، وهذا العدد من وجهة نظر مصممي HotBot كافٍ للأسباب التالية:

1- إن لم تكن النتائج الألف المعروضة مناسبة، فإن الطلب غير محدد تحديداً كافياً، ويجب أن يعيد المستخدم صياغته وتحديد المعلومات التي يريدها تحديداً دقيقاً.
2- إن عرض أكثر من 1000 نتيجة قد يبطّئ من فعالية البحث الذي يقوم به المستخدم، فعملية التجول بين هذه النتائج تأخذ زمناً طويلاً، لذا من الأفضل للمستخدم أن يعيد صياغة طلبه بدقة أكبر.

2. *البحث المتقدم*

يمكن القيام بالبحث المتقدم (Advanced Search) باستخدام HotBot، وتظهر واجهته في الشكل (6-23). ينبغي الإشارة إلى أن الإمكانات المتاحة ضمن صفحة البحث المتقدم هي إمكانات فلترة على النتائج، وليست بحثاً متقدماً حسب المفهوم الذي نستخدمه في نظم استرجاع المعلومات، أي بمعنى استخدام العوامل البوليانية (AND, OR, NOT) والعوامل المقاربة، والتي سنتحدث عنها بعد الحديث عن إمكانات الفلترة.

[3] سنتحدث عن طريقة إزعاج محركات البحث عندما نتحدث عن إضافة موقع إلى *HotBot*.



Look for وLanguage: موجودة في واجهة البحث ضمن الجزء 2 من الصفحة الرئيسية.

Word Filter: فلترة حسب الكلمات (ضمن الحقل الحواري).

Date: فلترة حسب التاريخ.

Page Must Include: فلترة حسب ما يجب أن تحويه الصفحة (صور، أو فيديو...).

Location/Domain: فلترة حسب الموقع الجغرافي (صفحات ضمن مواقع في أمريكا الشمالية، أو في الشرق الأوسط، أو...)، أو حسب المجال الذي ينتمي إليه الموقع

الشكل (6–23): صفحة البحث المتقدم في HotBot.

Word Stemming: استخدام المرادفات المختلفة لكلمات البحث.

Return Results: موجودة في واجهة البحث ضمن الجزء 2 من الصفحة الرئيسية.

3. *البحث المتقدم البولياني*

يسمح HotBot باستخدام العوامل البوليانية في صياغة طلبات البحث، هذه العوامل هي:

AND, OR, NOT. تسمح هذه العوامل بصياغة طلبات مركبة مثل:

(tourism AND Syria) AND (NOT Aleppo)

من أجل البحث عن صفحات تتحدث عن السياحة في سورية عدا تلك التي تتحدث عن حلب.

كل ما على المستخدم فعله من أجل صياغة طلبات كهذه هو تحديد الخيار Boolean phrase ضمن المنزلقة الأولى Look for.

4. لا يسمح HotBot باستخدام البحث المتقارب.

5. يسمح HotBot بالفلترة حسب مجموعة كبيرة من الوسائط (Parameters)، مثل التاريخ ولغة الصفحات والموقع الجغرافي، و... كما ذكرنا سابقاً.

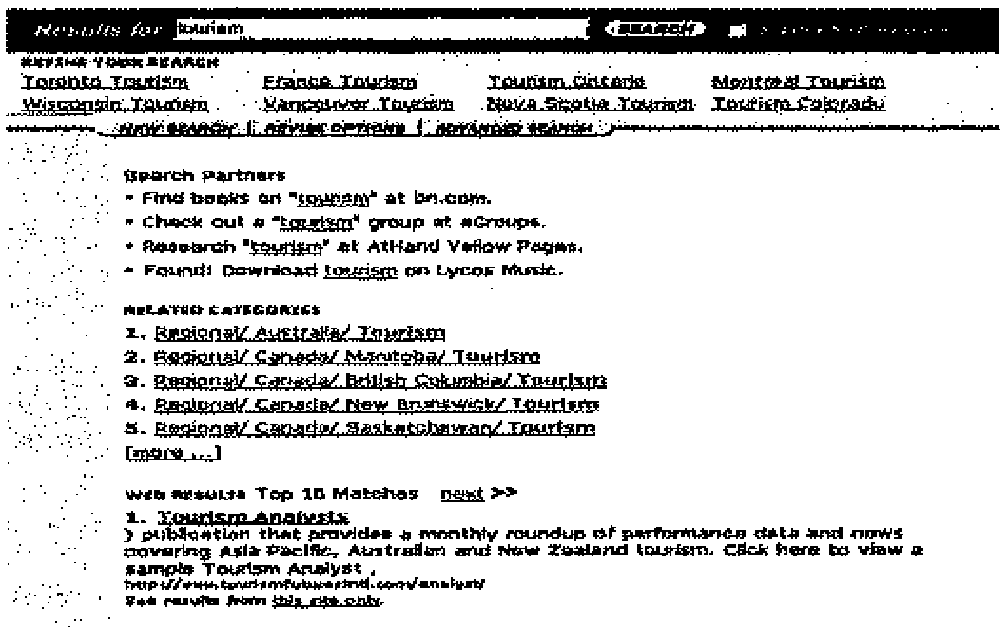
6. يؤمن HotBot تصنيفاً جيداً للمواضيع ضمن فئات بفضل مجلده المبني يدوياً، فنجد في الصفحة الرئيسية مجموعة من الفئات مثل: الثقافة والعلوم والفنون والألعاب وغيرها...يمكن للمستخدم التجول ضمن هذه الفئات، وضمن الفئات الفرعية التي تنبثق عنها، مما يحدد مجالات البحث، وبالتالي يفيد في تسريع الوصول إلى المعلومات.

7. *شكل النتائج*

ترتب النتائج في صفحات النتائج كما في الشكل (6–24)، وفيما يلي شرحٌ للارتباطات التشعبية الموجودة، وشكل النتائج:

- Search within these results: تسمح بالقيام بعمليات بحث جديدة على مجموعة النتائج الحالية.
- Refine Your Search: يقوم HotBot بتعديل طلب المستخدم الأساسي من تلقاء نفسه حسب معلومات خاصة لديه، ويعطي المستخدم إمكانية البحث باستخدام أحد الطلبات الجديدة. تظهر هذه الطلبات الجديدة تحت العنوان Refine Your Search، فمثلاً الارتباط التشعبي الذي يظهر على الشكل (6–24) بعنوان France tourism هو طلب بحث: +tourism france، أما Toronto tourism فهو طلب بحث: toronto +tourism.

- ☐ New Search : تسمح للمستخدم بإلغاء عملية البحث الحالية، والعودة إلى الصفحة الرئيسية من أجل صياغة طلب جديد.
- ☐ Revise Search : تسمح بتعديل خيارات البحث الذي قام المستخدم به.
- ☐ Advanced Search : تقوم بنقل الطلب إلى صفحة البحث المتقدم، حيث تمكن إضافة مجموعة جديدة من خيارات الفلترة على نتائج الطلب.
- ☐ Search Partners : مجموعة من الأدوات التي يمكن استخدامها في البحث.



الشكل (6–24): صفحة النتائج

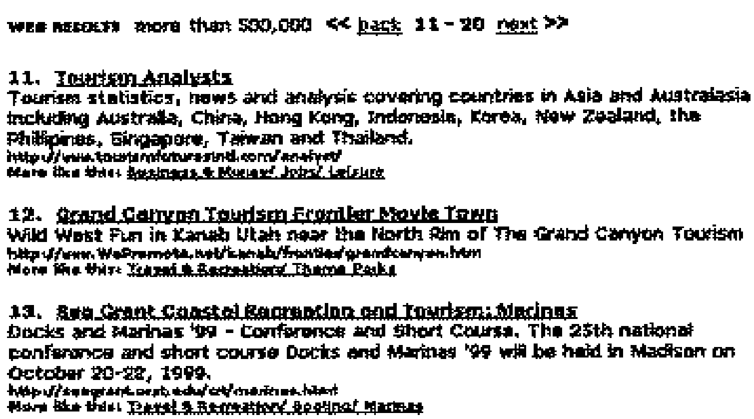
أما نتائج البحث فهي مصنّفة كما يلي:

- ☐ Related Categories : الفئات المتعلقة بموضوع البحث من مجلد HotBot.

- Top 10 Results: مجموعة المواقع التي أعطاها HotBot أعلى درجات مطابقة. يُحدد HotBot هذه المواقع حسب مجموعة من المعايير مثل المواقع التي يزورها المستخدمون عادةً نتيجة عمليات البحث، والتصنيف العام للمواقع بعضها بالنسبة لبعض.

- This site only: في الحالات التي يكون فيها عدد الصفحات النتائج من موقع معين كبيراً، يكتفي HotBot بارتباط تشعبي إلى الصفحة الأساسية للموقع (this site only) بدلاً من عرض كامل صفحات الموقع المطابقة. إن هذه الطريقة معروفة في نظم استرجاع المعلومات باسم التجميع Clustering.

- Main Results: لا تظهر هذه النتائج على نفس الصفحة التي تظهر عليها الإجابات العشر الأولى Top 10 Results، إذ يجب الانتقال إلى صفحة النتائج التالية من أجل مشاهدة هذه النتائج، حيث نجد في أعلاها عدد نتائج البحث التي وجدها HotBot، كما يُظهر ذلك الشكل (25-6).



الشكل (6-25): صفحة النتائج التي تظهر فيها Main results.

تختلف طريقة عرض Top 10 Results عن طريقة عرض Main Results لأن طرق ترتيب النتائج مختلفة، فلا يمكن إعطاء نسب مئوية للمطابقة من أجل Top 10 results.

نلاحظ وجود تاريخ بجانب URL النتائج، وهو يدل على تاريخ إضافة الصفحة (أو الفحص الأخير لها) إلى قاعدة بيانات HotBot.

8. يسمح HotBot بالتغذية الراجعة، فهو يتيح للمستخدم أن يقوم بعمليات بحث جديدة على مجموعة النتائج التي اقترحها، وذلك عن طريق الخيار Search within these results في أعلى صفحة النتائج (راجع الشكل (6–24)).

9. يتمتع HotBot بسرعة جيدة، فلا نعاني من الانتظار من أجل الحصول على النتائج، رغم ضخامة قاعدة بياناته.

10. معايير أخرى

يؤمن HotBot مجموعة من Meta Words تسمح بالبحث عن معلومات غير نصية انطلاقاً من واجهة البحث النصية. وهي عبارة عن زوج من الكلمات: *كلمة مفتاحية:قيمة*، يمكن الاستفادة منها كما يلي: استخدام Meta Word مثل "title:president" ضمن طلب ما يعني أن الصفحات العادة ستحوي كلمة president ضمن عناوينها.

Meta Word	الشرح
Domain:[name]	يخصص البحث ضمن المجال المحدد. يمكن للمجالات أن تحدد بعمق ثلاثة مستويات فقط، مثلاً: support.intel.com, intel.com, com
Depth:[number]	يهمل الصفحات التي عمقها أكبر من number
Feature:acrobat	يكتشف ملفات Acrobat
Feature:applet	يكتشف بريمجات جافا

Feature:activex	يكتشف ActiveX Controls
Feature:audio	يكتشف مجموعة من أنماط الملفات الصوتية
Feature:embed	يكتشف plug-ins
Feature:flash	يكتشف Flash plug-ins ضمن HTML
Feature:form	يكتشف استخدام Forms ضمن HTML
Feature:frame	يكتشف استخدام Frames ضمن HTML
Feature:image	يكتشف ملفات صور (GIF, JPEG, etc.)
Feature:script	يكتشف استخدام Script
Feature:shockwave	يكتشف ملفات Shockwave
Feature:table	يكتشف استخدام جداول ضمن HTML
Feature:video	يكتشف ملفات الفيديو من أنماط معينة
Feature:vrml	يكتشف ملفات VRML

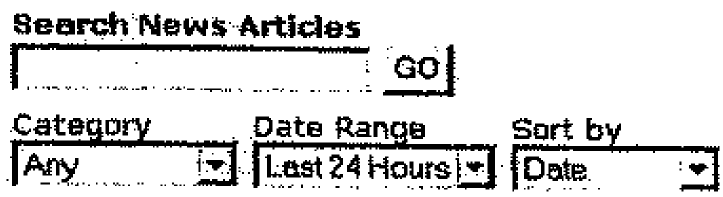
الجدول (6-1): الـ **Meta Words** التي يستخدمها **HotBot**

ج. البحث حسب الموصفات

يتيح HotBot خدمات يمكن البحث فيها بطلبات من نمط تلك التي نستخدمها في قواعد البيانات، مثل البحث عن الأشخاص، والصفحات الصفر، وسنتكلم عنها لاحقاً في خدمات المعلومات الأخرى.

6–2–3 خدمات المعلومات الأخرى

1. الأداة الأولى هي News Headlines، وهي تسمح بعرض مجموعة من الأخبار اليومية، كما تتيح البحث عن أي مقالات إخبارية يطلبها المستخدم كما في الشكل (6–26).



الشكل (6–26): واجهة البحث عن المقالات الإخبارية في HotBot

نلاحظ أن واجهة البحث هذه تتيح فلترة المقالات الناتجة باستخدام فئة الموضوع (Category) (أي هل المقال سياسي أم علمي أم ثقافي أم...)، أو باستخدام مجال لتاريخ النشر (Date Range) (مثل 6 ساعات، 24 ساعة، أسبوع، شهر). كما تتيح إمكانية فرز المقالات الناتجة حسب تاريخ النشر أو حسب مطابقتها لطلب المستخدم.

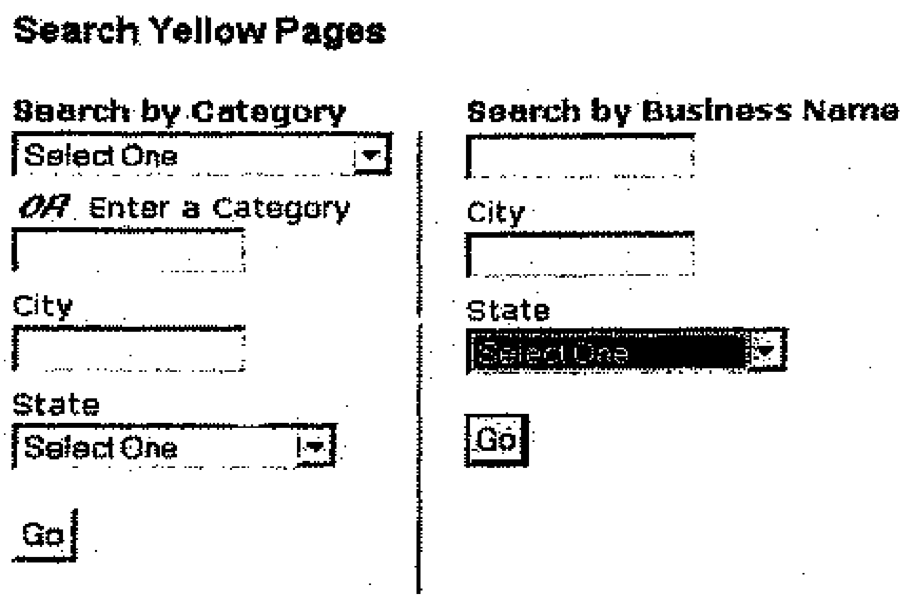
2. الأداة الثانية هي Email Addresses.

3. الأداة الثالثة هي Stock Quotes.

4. الأداة الرابعة هي Discussion Groups، وتسمح بإجراء عمليات بحث ضمن Usenet عن سؤال أو موضوع معين.

5. الأداة الخامسة هي White Pages.

6. الأداة السادسة هي Yellow Pages، وتسمح بالبحث في الأوراق الصفراء، وهي صفحات تحوي معلومات عن النشاط الاقتصادي في مناطق معينة. يمكن البحث في الأوراق الصفراء ضمن الولايات المتحدة حصراً باستخدام واجهة البحث المعروضة في الشكل (6–27).



الشكل (6–27): واجهة البحث ضمن الصفحات الصفر في HotBot

6–2–4 إضافة موقع إلى محرك البحث HotBot

يمكن إضافة موقع إلى HotBot بنفس الطريقة التي تحدثنا عنها في !Yahoo، أي التجول بين فئـات المجلد حتى الوصول إلى الفئة المناسبة، ثم تسجيل الموقع فيها. يمكن أيضاً إضافة URL إلى HotBot من خلال صفحته الرئيسية، من خلال ضغط الزر Add URL بين الجزأين [2] و[3].

يحتاج HotBot حتى يسجل الصفحة الجديدة من 3 أسابيع إلى 60 يوماً كحد أقصى، وهو الزمن اللازم حتى يتمكن عنكبوته الخاص (Spider) من مسح مواقع الشبكة.

يجب أن يُسجِّل المستخدم كامل صفحات موقعه ضمن HotBot، على ألا يزيد عدد الصفحات المسجلة من نفس الموقع على 50 صفحة في كل يوم. سيقوم المستخدم بهذه العملية إن كان مستعجلاً، أمّا إن كان الوقت غير مهم بالنسبة له، فلن يضطر إلى تسجيل كامل صفحات الموقع بنفسه، إذ يمكنه الانتظار على الأكثر 60 يوماً، وهو الزمن اللازم حتى يتمكن عنكبوت HotBot من زيارة موقع

المستخدم وفهرسة صفحاته بنفسه. يزور HotBot المواقع التي يعرف عناوينها فقط، لذا يجب على المستخدم أن يسجل عنوان موقعه فقط، والباقي يتولاّه HotBot.

قد يحتاج المستخدم في بعض الأحيان إلى منع HotBot من فهرسة مجموعة من صفحات موقعه، مثل الصفحات التي تحوي معلومات شخصية خاصة بالمستخدم، ولا يريد لأحد غريب أن يطلع عليها. لمنع HotBot من فهرسة صفحة ما، يمكن تحديد مجموعة المجلدات و الملفات التي لا يريد المستخدم إضافتها إلى قاعدة بيانات HotBot ضمن ملف خاص على موقعه باسم Robots.txt. سيقوم عنكبوت HotBot بقراءة هذا الملف وتنفيذ الأوامر التي فيه (من أجل قواعد كتابة الأوامر ضمن هذا الملف يمكن زيارة الموقع التالي:

(http://www.hotwired.com/webmonkey/99/02/index3a_page4.html?tw=javascript.

توجد طريقة أخرى لمنع HotBot من فهرسة الصفحة باستعمال مؤثر HTML التالي:

<META NAME="ROBOTS" CONTENT="NOINDEX">

يقوم HotBot بفهرسة الصفحة فهرسةً كاملة، لكنه يسمح للمستخدم بأن يُحدد مجموعة من الكلمات المفتاحية الخاصة بصفحته، وتحديد شرح عن هذه الصفحة (يستخدمه HotBot أثناء إظهار النتائج) باستخدام مؤثرات HTML التالية:

<META NAME="author" CONTENT="***your name goes here***">
<META NAME="description"CONTENT="***your site description goes here***">
<META NAME="keywords" CONTENT="***keywords for your site go here***">

يمكن للمستخدم أن يضيف المعلومات التي يريدها مكان النص المكتوب بالمائل والداكن.

إن الكلمات المفتاحية التي يُضيفها المستخدم في مؤثر keywords لها تأثير على ترتيب الصفحة ضمن نتائج البحث كما ذكرنا عند شرح طريقة ترتيب النتائج. يعمد البعض إلى تحديد مجموعة كلمات مفتاحية ضمن مؤثرات HTML الخاصة بذلك دون أن تمتّ للموضوع بصلة، أو قد يعمدون إلى تكرار كلمات مفتاحية معينة بغية إعطاء الصفحة درجة مطابقة عالية، والهدف هو إزعاج محركات البحث أو إعطاء صفحتهم درجة ترتيب عالية ضمن نتائج البحث.

تُعتبر هذه العملية سيئة لمحرك البحث والمستخدمين على السواء، فهي توقع محرك البحث في أخطاء عند ترتيب الصفحات، كما أنها تسيء إلى المستخدم الذي يقوم بالبحث لأنه لن يتمكن من العثور على طلبه من المعلومات. يستخدم HotBot مجموعة من المعايير لتحديد فيما إذا كانت الصفحةُ صفحةَ إزعاج أم لا، كما أنه يتعرف على المواقع التي تعمد إلى إضافة صفحات إزعاج-ويعاقبها بأن يخفض درجة ترتيب الصفحات التي يجدها ضمن تلك المواقع، لذا يجب على المستخدم الحريص ألاّ يقوم بأعمال كهذه وإلا سينقلب الأمر عليه.

يضع HotBot حداً على طول الكلمات المفتاحية والشروح ضمن هذه المؤشرات، فيسمح بـ 75 حرفاً للكلمات المفتاحية و150 حرفاً للشرح من أجل الصفحة الواحدة.

6-3 محرك البحث AltaVista

WWW.AltaVista.com

يُعد AltaVista أحد أكبر محركات البحث على الإنترنت، فنهرسه يتسع لأكثر من 150 مليون صفحة وب، لكن إجاباته تعطي الكثير من الضجيج (صفحات غير ملائمة لطلب المستخدم). تكمن قوته الأساسية في إنجاز الطلبات المخصصة، لأنه يتيح خيارات كثيرة في صياغة الطلب.

يُنصح عند استخدام AltaVista في البحث بأن يصيغ المستخدم طلباته باللغة الإنكليزية. فمثلاً إذا أراد المستخدم البحث عن مواقع تتحدث عن السياحة في سورية، فمن الأفضل أن يصيغ طلبه بالشكل: "I want information about tourism in syria".

كما يفيد استخدام جملة (بين "") ضمن الطلب في تحسين نتائج البحث تحسيناً جيداً. يُنصح أيضاً باستخدام إمكانيات الفلترة (مثلاً الفلترة حسب لغة الصفحة)، والانتباه إلى أن AltaVista يميز بين الأحرف الصغيرة والكبيرة، لذا من الأفضل في الحالات العامة أن نستخدم الأحرف الصغيرة.

هنالك ملاحظة هامة بخصوص النتائج (الإجابات) التي يعيدها AltaVista، وهي اختلاف عدد هذه النتائج لنفس الطلب من وقت لآخر. يعود السبب إلى التقنيات التي يستخدمها AltaVista في الحصول على هذه النتائج. فقد يختلف التقدير عندما ينتقل المستخدم من صفحة لأخرى، كما يختلف حسب الوقت من النهار، إذ يقوم AltVista بتحديد زمن معين لكل طلب بحث أثناء ضغط العمل عليه. لذلك ينصح AltaVista مستخدميه بالعمل خلال الفترة 3AM, 1PM GMT.

يعيد AltaVista 200 نتيجة فقط كحد أعلى لكل طلب يقوم به المستخدم، لئلا يتسبب ذلك في ازدحام خطوط النقل. أمّا إذا أراد المستخدم الحصول على عدد أكبر من النتائج، فيمكنه عندئذ القيام ببحث متقدم، ويحصل في هذه الحالة على آلاف النتائج، دون مراعاة الازدحام على خطوط الشبكة.

6-3-1 واجهة AltaVista

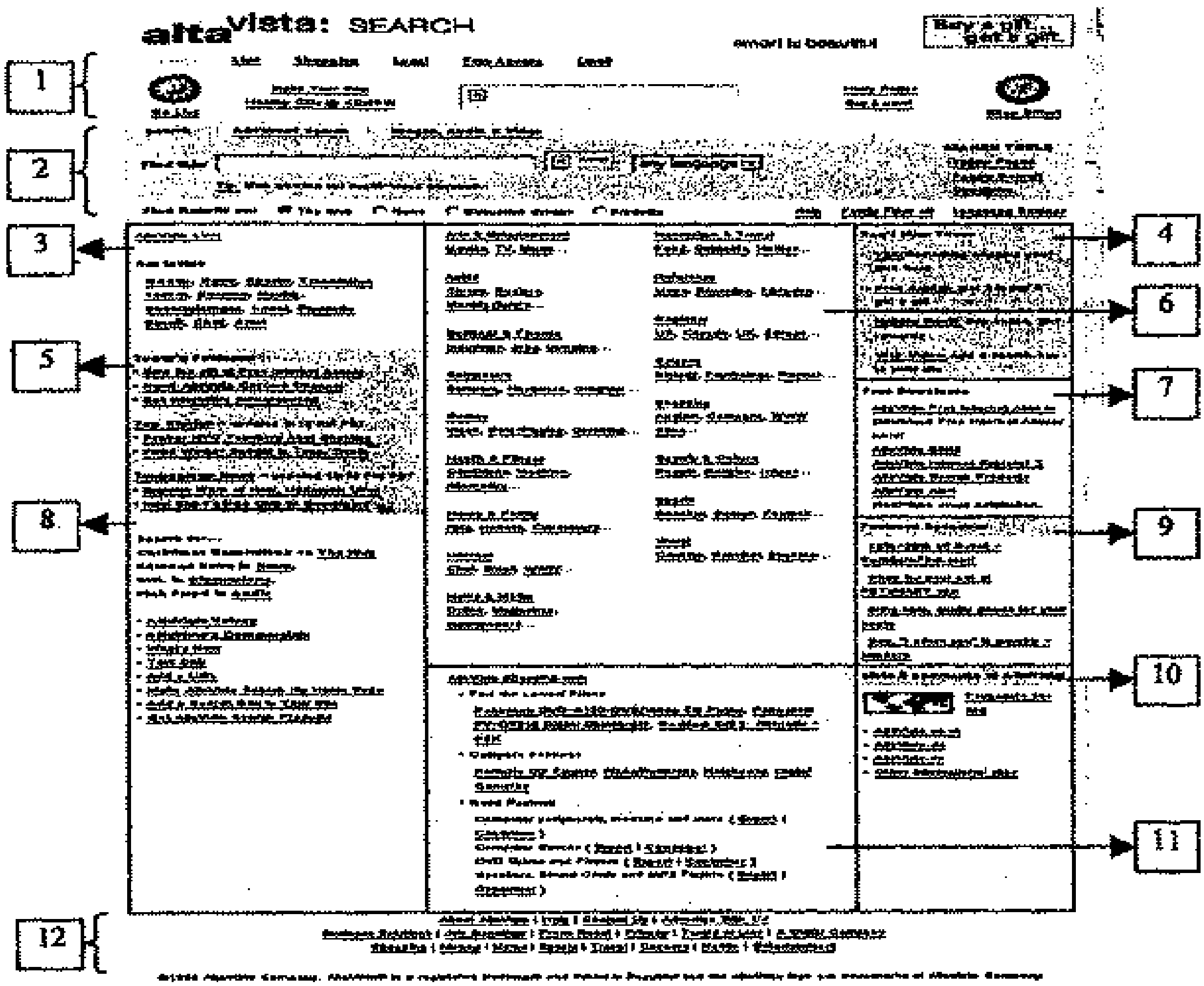
يُظهر الشكل (6-28) صفحة Alta Vista على الوب، والتي يمكن الوصول إليها باستخدام العنوان: http://www.altavista.com، وهي واجهة ضخمة جداً، يمكن للمستخدم أن يصل إلى الكثير من المعلومات من خلالها. تحتوي هذه الصفحة جميع إمكانات البحث التي يقدّمها AltaVista والعديد من خدمات المعلومات الأخرى. نجد فيما يلي الإمكانات المختلفة لـ AltaVista:

* الجزء [1]: يتضمن مجموعة من الدعايات والإعلانات التجارية، إضافةً إلى مجموعة من الارتباطات التشعبية إلى الأدوات التي يقدمها AltaVista مثل: Search, Email, Shopping.
* الجزء [2]: يسمح للمستخدم بالقيام بعمليات بحث.
* الجزء [3]: يتضمن مجموعة من خدمات AltaVista.

- الجزء [4]: يتضمن مجموعة أخرى من الخدمات.
- الجزء [5]: يتضمن مجموعة من التقنيات والأخبار الجديدة.
- الجزء [6]: يسمح للمستخدم بالتصفح ضمن مجلد AltaVista.
- الجزء [7]: يتضمن مجموعة من البرامج التي يستطيع المستخدم تحميلها مجاناً على جهازه.
- الجزء [8]: يتضمن مواضيع مقترحة لعمليات البحث.
- الجزء [9]: يتضمن مجموعة من العروضات للتسوق.
- الجزء [10]: يسمح للمستخدم بزيارة مواقع AltaVista في بلدان أخرى.
- الجزء [11]: يسمح للمستخدم بالقيام بعمليات تسوق.
- الجزء [12]: يتضمن مجموعة من الارتباطات التشعبية إلى بعض خدمات AltaVista.

6-3-2 البحث عن المعلومات في AltaVista

يتيح AltaVista الأنواع الثلاثة من طرق البحث عن المعلومات: التصفح، البحث عن المحتوى، البحث عن الموصفات.



الشكل (6–28): صفحة AltaVista

آ. التصفح

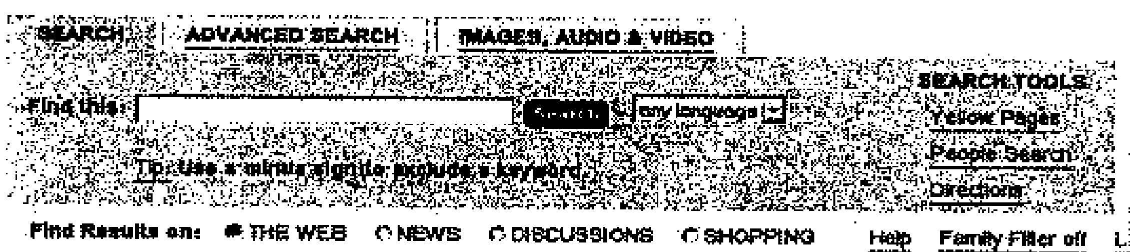
يشابه البحث بالتصفح في AltaVista التصفح في باقي محركات البحث، لكن AltaVista لا يملك مجلداً خاصاً به، بل يستفيد من مجلد LookSmart من أجل تقديم هذه الخدمة للمستخدمين. ويمكن التصفح ضمنه باستخدام الارتباطات التشعبية ضمن الجزء 6 ، وسنسمي هذا المجلد هنا تجاوزاً مجلد AltaVista.

يملك مجلد AltaVista 17 فئة رئيسية مختلفة تظهر على الشكل (6–28)، تتيح هذه الفئات إمكانية البحث بالتصفح. تتفرع هذه الفئات الأساسية إلى مجموعة من الفئات الفرعية التي يمكن الوصول إليها باتباع الارتباطات التشعبية. ويُذكر أن AltaVista سيشترك في مشروع المجلد المفتوح[4] (Open Directory) قريباً.

ب. البحث عن المحتوى

يتيح AltaVista إمكانية صياغة طلبات البحث في الجزء 2 من صفحته الرئيسية. يُظهر الشكل (6–29) واجهة البحث هذه، وسنعرض مزاياها المختلفة فيما يلي:

- الحقل الحواري هو مكان صياغة طلبات المستخدم.
- Tip: يقود إلى مجموعة من الأمثلة والمعلومات المفيدة عن طريقة القيام بالبحث ضمن AltaVista بفاعلية أكبر. تتبدل هذه الأمثلة والمعلومات تبدلاً مستمراً، لذا ننصح بإلقاء نظرة عليها من وقت لآخر.
- منزلقة تحديد اللغة (ويظهر ضمنها النص any language)، وتفيد في فلترة الصفحات التي يجدها النظام مطابقة لطلب المستخدم، فلا ينتج عن البحث إلا الصفحات المكتوبة باللغة المحددة. يدعم AltaVista حتى الآن 25 لغة فقط، واللغة العربية ليست منها.



الشكل (6–29): واجهة البحث ضمن الصفحة الرئيسية لـ AltaVista

---

4 كنّا قد تحدثنا عن المجلد المفتوح عند الحديث عن محرك البحث *HotBot*.

☐ مجموعة الخيارات أسفل الشكل (عنوانها Find Results on)، وتفيد في تحديد منطقة البحث. الخيار الافتراضي (يظهر في الشكل) هو البحث في كامل الوب، ويمكن تغيير الخيار ليقتصر البحث على الأخبار News، أو مجموعات الحوار Discussion Groups، أو ...

☐ مجموعة من الارتباطات إلى:

- معلومات مساعدة عن الصفحة الحالية (Help).
- صفحة ضبط وإعداد Family Filter وسنشرحها في "خدمات المعلومات الأخرى".
- ضبط إعدادات AltaVista على لغة معينة Language Settings (وهي غير ظاهرة على الشكل، لكنها تقع على نفس السطر الذي يوجد عليه الارتباطان السابقان).

في أعلى الشكل، نلاحظ ثلاثة تبويبات هي:

Search | Advanced Search | Images, Audio &Video. Search هو التبويب الافتراضي، وAdvanced Search هو تبويب يسمح بإجراء عمليات بحث أكثر تعقيداً، أما التبويب الأخير فيسمح بالبحث عن وسائط متعددة غير نصية على الوب. لصفحة Images, Audio &Video الشكل الموضح في الشكل (6–30).



الشكل (6–30): واجهة البحث عن الوسائط المتعددة غير النصية في AltaVista

- Find results on: تحدّد نوع الوسائط التي نبحث عنها، صورة أم صوت أم فيديو.
- Tip: أمثلة ومعلومات مفيدة للقيام ببحث أكثر فاعلية.
- From: منطقة البحث: الوب و مجموعات AltaVista.

- Limit to one result per URL : يعطي صفحة مطابقة واحدة فقط من كل موقع.
- مجموعة من الارتباطات التشعبية إلى Help, Family Filter, Language Settings.

تقييم البحث بالمحتوى

1. البحث البسيط

يسمح AltaVista بعمليات البحث البسيط، ومن أجل فحص إمكانات هذا النوع من البحث، قمنا بالتجارب التالية:

- التحسس للأحرف الصغيرة والكبيرة: بالبحث عن tourism و Tourism، نجد اختلافاً في عدد صفحات المعادة في كل مرة. مما يعني أنه يتحسس للأحرف الكبيرة والصغيرة. يتحسس AltaVista للأحرف الصغيرة والكبيرة على النحو التالي: إن وُجد حرف كبير ضمن الكلمة، فإن البحث يصبح حساساً للأحرف الصغيرة والكبيرة، أما إن كانت جميع الأحرف صغيرة فإن البحث يصبح غير حساس للأحرف الكبيرة والصغيرة. بمعنى آخر، البحث عن tourism يقبل كلمات مثل TOURISM أو Tourism، أما البحث عن Tourism فلا يقبل tourism أو TOURISM أو TouRism، بل يقبل Tourism فقط.
- البحث عن جمل: بالبحث عن "tourism in Syria"، نجد أن الصفحات التي وجدها النظام تتحدث عن السياحة في سورية، وتحوي هذه الجملة. إذن يتيح إمكانية تجميع كلمات البحث في جمل، إذ يكفي وضع الجملة ضمن "" حتى يقوم بعملية بحث عن الصفحات التي تحوي الجملة بكاملها. يمكن تحديد جملة بطريقة أخرى، وذلك باستخدام مجموعة من الرموز الخاصة مثل _,$,/,#,%، فالبحث عن Syrian#airlines مطابق للبحث عن "Syrian airlines".

- الكلمات المفتاحية اللازمة: يسمح AltaVista للمستخدم بأن يحدد كلمة يريدها أن تكون محتواة ضمن جميع الصفحات المطابقة. فمثلاً يؤدي الطلب: tourism +Syria إلى إيجاد الصفحات التي تتحدث عن tourism وتحتوي ضمنها بالضرورة كلمة Syria.
- الكلمات المفتاحية الممنوعة: يتيح AltaVista إمكانية منع كلمة من أن تكون موجودة ضمن صفحة ما عن طريق استخدام "-". فمثلاً: بالبحث عن "tourism" والبحث عن "tourism -Syria"، نجد بفحص النتائج أن الصفحات التي تتحدث عن "Syria" قد اختفت في الطلب الثاني.
- يسمح باستخدام الرمز * في آخر الكلمة فقط من أجل الدلالة على عدد غير محدّد من الأحرف، كما في المثال: "*Syr"، حيث يبحث عن جميع الصفحات التي تحوي كلمات تبدأ بـ "Syr".
- يرتب AltaVista النتائج حسب درجة مطابقتها، ويحسب درجة المطابقة حسب خوارزمية خاصة تأخذ بعين الاعتبار العوامل التالية:
  - ☐ عدد كلمات البحث التي تحويها الصفحة.
  - ☐ قرب كلمات البحث بعضها من بعض ضمن الصفحة.
  - ☐ تُحذف صفحات المواقع التي تقوم بعمليات خداع لمحرك البحث[5] من النتائج.

2. البحث المتقدم

يسمح AltaVista باستخدام البحث المتقدم بنوعيه: البحث البولياني والبحث المقارب.

يعرض الشكل (6-31) واجهة البحث المتقدم ونجد فيه:

☐ Boolean query: حقل الحوار، يسمح بكتابة الطلبات البوليانية فقط (تحوي عوامل بوليانية).

---

[5] كنا قد شرحنا مسألة خداع محركات البحث أثناء الحديث عن *HotBot*، ويستخدم *AltaVista* برمجيات خاصة من أجل كشفه.

- Sort by : لتحديد الكلمات (من بين كلمات البحث) التي يريد المستخدم أن تُرتَّب النتائج وفقاً لها. إن عدم إدخال أي كلمة ضمن هذا الحقل يعني أن نتائج البحث لن تكون مرتبة.
- Language : فلترة النتائج حسب لغة الصفحات.
- حقول التاريخ: تسمح بفلترة النتائج حسب التاريخ. يمكن تحديد تاريخ بداية وتاريخ نهاية، وفق الترتيب: DAY/MONTH/YEAR.

يمكن القيام بالبحث البولياني باستخدام العوامل البوليانية التي يؤمنها AltaVista وهي: AND, OR, AND NOT (لا يسمح بـ NOT لوحدها).

يسمح AltaVista باستخدام الرموز التالية من أجل العوامل البوليانية:

&: من أجل العامل AND.

|: من أجل العامل OR.

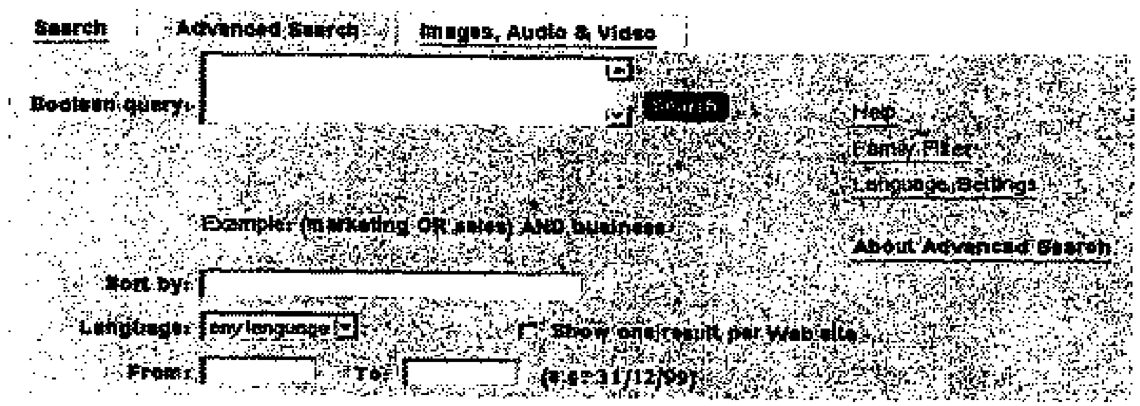
!: من أجل العامل AND NOT.

وعليه، يمكننا أن نصوغ الطلب:

(tourism & Syria) ! Aleppo

أو (tourism AND Syria) AND NOT Aleppo

للبحث عن الصفحات التي تتحدث عن السياحة في سورية عدا تلك التي تتحدث عن حلب.



الشكل (6-31): البحث المتقدم في AltaVista

يمكن القيام بالبحث المتقارب باستخدام التابع NEAR (يستخدم AltaVista الرمز ~ أيضاً للعملية نفسها). يسمح استخدام العامل NEAR في الطلب (مثل tourism NEAR Syria) باستحضار الصفحات التي تحوي الكلمتين tourism وSyria على أن لا يكون عدد الكلمات الأخرى بين هاتين الكلمتين أكثر من 10 كلمات.

يجب الانتباه إلى أن صياغة الطلب في البحث المتقدم تختلف عنه في البحث العادي، فيجب استخدام العوامل البوليانية ضمن طلب البحث المتقدم، ولا يمكن استخدام "" (البحث عن جملة) ضمنه، كما لا يمكن أن توجد كلمتان دون تحديد عامل بولياني بينهما.

3. الفلترة

يسمح AltaVista بالفلترة حسب لغة الصفحات، وحسب مجال زمنيّ، وحسب مضمون الصفحات (باستخدام Family Filter).

4. يؤمن AltaVista تصنيفاً جيداً للمواضيع ضمن فئات بفضل مجلده المبني يدوياً، ونجد الفئات الرئيسية لمجلده ضمن الجزء 6 من الصفحة الرئيسية.

5. إظهار النتائج

تظهر نتائج AltaVista ضمن ثمانية أقسام مستقلة بعضها عن بعض، ويمكن الانتقال من أحدها إلى الآخر، كما نلاحظ في الشكل (6–32).

يبحث AltaVista ضمن مجموعة من مصادر المعلومات المختلفة وهي: الصور، والصوت، والفيديو، ومجلده الخاص، وصفحات الوب، والأخبار، ومجموعات الحوار، والتسوق، ويقودنا هذا إلى الأقسام الثمانية التي تُقسم إليها النتائج.

Products	News	Discussion Groups	The Web
Images	Audio	Video	Categories

الشكل (6–32): أقسام نتائج البحث ضمن AltaVista

– صفحات الوب

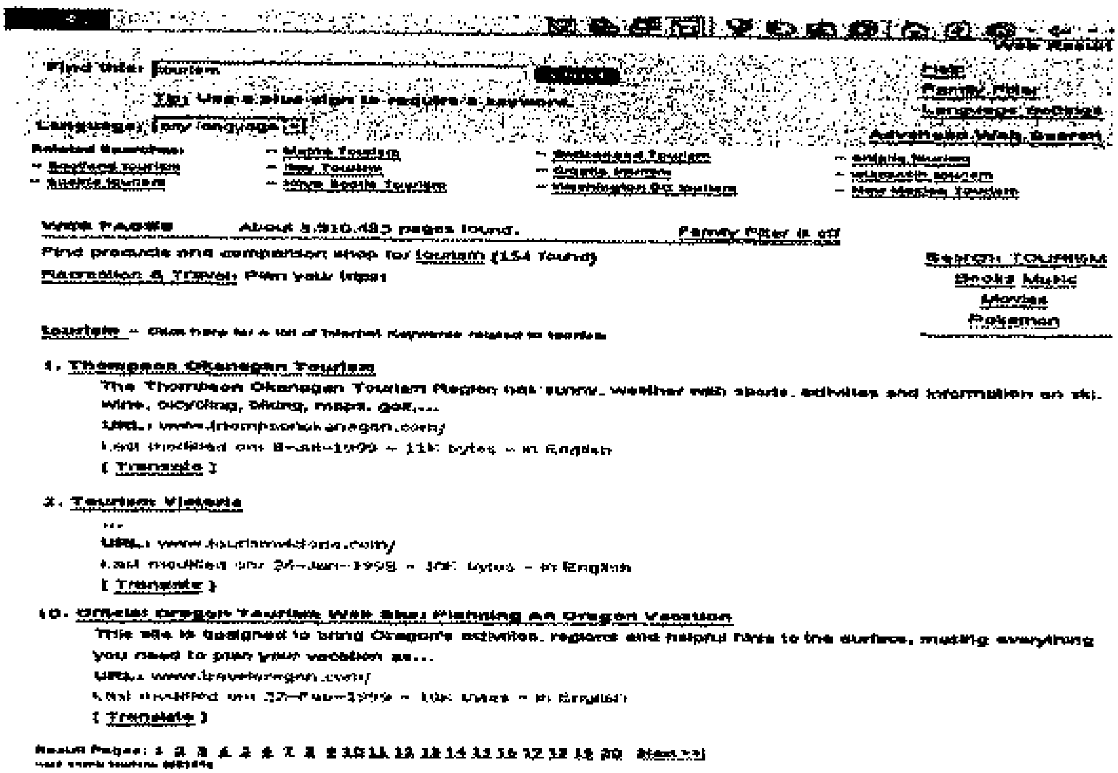
وهي مجموعة الصفحات التي وجدها AltaVista مطابقة لطلب المستخدم. نلاحظ في الشكل (6–33) طريقة عرض النتائج، وهي معدّلة كي تُظهر الأمور الأساسية.

أعلى الصفحة، وتحت حقل الحوار، يوجد جزء بعنوان Related Searches، وهي مجموعة من طلبات البحث التي يصيغها AltaVista بنفسه انطلاقاً من طلب المستخدم. فمثلاً، نجد على الشكل أن الارتباط Scotland tourism هو الطلب Scotland +tourism، وباتباع الارتباط التشعبي، فإننا نقوم بتنفيذ هذا الطلب. يظهر هذا الجزء من أجل صفحات الوب فقط، ولا يظهر في باقي الأقسام.

أسفل مجموعة الارتباطات هذه، نرى عدد النتائج التي وجدها AltaVista وهو هنا 3310485.

تظهر الصفحات مرتبة حسب درجة مطابقتها لطلب المستخدم، ونجد عشر نتائج فقط. يظهر في كل نتيجة عنوان الصفحة، وشرح بسيط عنها (سنشرح كيفية إضافة الشرح إلى الصفحة عند الحديث عن

إضافة موقع إلى AltaVista). أسفل كل شرح، يوجد URL هذه الصفحة، وتاريخ آخر تعديـل لهـا، وحجم ملفها، وارتباطٌ إلى أداة الترجمة في AltaVista وهي BABELFISH (سنتحدث عنها لاحقاً).



الشكل (6-33): نتائج البحث في AltaVista ضمن قسم صفحات الويب.

توجد في أسفل الصفحة ارتباطات إلى باقي صفحات النتائج. عدد الصفحات كما نرى على الشكل 20 صفحة، وكل صفحة تحوي 10 نتائج، إذن عدد النتـائج المعادة هـو 200 نتيجـة فقـط، بينما عـدد النتائج المتوقعة الظاهر في أعلى الصفحة هو 3310485 صفحة.

ب. الفئات

هي مجموعة الفئات المطابقة (الشكل 6-34) لطلب المستخدم والتي وجدها AltaVista ضمن مجلده. لم تُظهر في الشكل صفحة النتائج كاملة، فأعلى صفحة النتائج له نفس شكل صفحة نتائج قسم صفحات الوب. نلاحظ في الشكل أن AltaVista يعطي عدد الفئات المطابقة لطلب المستخدم في الأعلى، وهو هنا 2462 فئة. تظهر الفئات النتائج مع جميع الفئات الآباء لها، حتى نصل إلى إحدى الفئات الرئيسية الموجودة في الصفحة الرئيسية ضمن الجزء 6. تظهر في كل صفحة نتائج 10 نتائج فقط، وفي الشكل (6-35)، أظهرنا خمس نتائج فقط، وفي أسفل الصفحة ارتباطات إلى باقي صفحات النتائج، وعددها 20 صفحة.

CATEGORY RESULTS 2,462 categories found.

Find products and comparison shop for tourism (154 found)

Recreation & Travel: Plan your trips!

Top / Business / Training and Schools / Travel and Tourism *(5)*

Top / Business / Marketing / Consultants / Tourism *(8)*

Top / Regional / North America / Mexico / Travel and Tourism *(299)*

Top / Regional / North America / Mexico / Travel and Tourism / Restaurants *(0)*

Top / Regional / North America / Mexico / Travel and Tourism / Travel Agents *(4)*

Result Pages: 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 [Next >>]

الشكل (6-35) : نتائج البحث في AltaVista ضمن قسم الفئات

تدل الأرقام بجانب كل فئة على عدد المواقع المطابقة ضمن الفئة.

ج. الصوت

كما نلاحظ في الشكل (6-36)، يظهر أولاً عدد النتائج التي وجدها AltaVista (27)، ثم تظهر كل نتيجة كما هو موضح: أولاً عنوان الصفحة التي تحوي الصوت، ثم URL الصفحة، ثم اسم ملف الصوت، ونمطه، وأخيراً مالك حقوق النشر.

AUDIO RESULTS About 27 audios found. Family Filter is on

Page Title Headhunter – Tourism

Page URL http://cdnow.com/switch/from=cr-4848614/...D-88111+11759+2

File Name ra.ram

File Type ra

Copyright (c) (P) 1998 MCA Records

more info

الشكل (6-36): نتائج البحث في AltaVista ضمن قسم الصوت.

لم تُظهر في الشكل سوى نتيجة واحدة، علماً أن صفحة النتائج الواحدة تتسع لعشر نتائج.

د. الأخبار

نلاحظ في الشكل (6-37) ظهور عدد النتائج التي وجدها AltaVista، ثم تظهر كل نتيجة كما هو موضح: أولاً عنوان المقال (الخبر)، وتحته موجز عنه، ثم مصدر الخبر وتاريخ صدوره. لم تُظهر في الشكل سوى نتيجة واحدة، علماً أن الصفحة تتسع لعشر نتائج.

NEWS RESULTS 77 stories found.

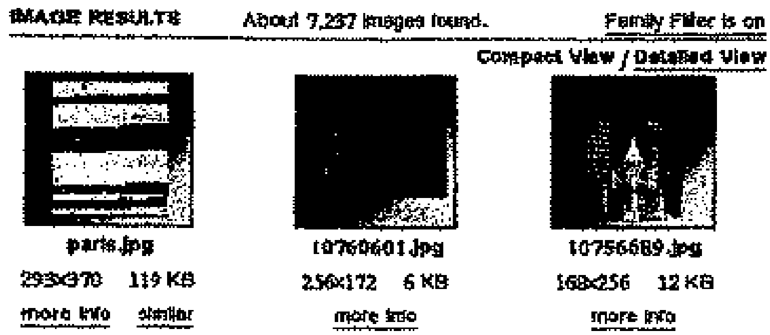
1. INTERVIEW–Morocco tourism set for recovery
RABAT, Dec 13 (Reuters) – Morocco's tourism industry is recovering rapidly and the government sees a 28 percent rise in the number of foreign visitors this year, compared to 1998. "We expect up to...
Source: Reuters 12/13/1999 10:06:00

الشكل (6-37): نتائج البحث في AltaVista ضمن قسم الأخبار.

هـ. الصور

تظهر في الشكل (6-38) طريقة عرض نتائج البحث ضمن قسم الصور. النتائج التي نراها ناتجة عن عملية البحث عن صور أبراج. نجد في أعلى الصفحة عدد النتائج التي وجدها النظام، ثم الصورة

ومعلومات عنها: اسم ملف الصورة، وأبعادها، وحجمها. تتسع صفحة النتائج الواحدة لـ 12 نتيجة، وقد عرضنا منها النتائج الثلاث الأولى فقط.



الشكل (6-38): نتائج البحث في AltaVista ضمن قسم الصور.

و. الفيديو

تُعرض النتائج في هذا القسم بنفس طريقة عرض نتائج الصور، حيث يَعرض AltaVista صورة عن الفيديو الذي يجده، وفي أسفله اسم ملف الفيديو، ونمطه، ومدة الفيلم، وزمن العرض.

ز. المنتجات

إن النتائج التي يجدها AltaVista ناتجة عن البحث ضمن مجلده، ونلاحظ أن النتائج هي فئات من المجلد، أما الرقم الذي يظهر بجانب الفئة، فهو عدد المواقع المطابقة للبحث ضمن هذه الفئة.

RELEVANT SHOPPING CATEGORIES　　Results found for "tourism": 221

Click into each category to see your results.

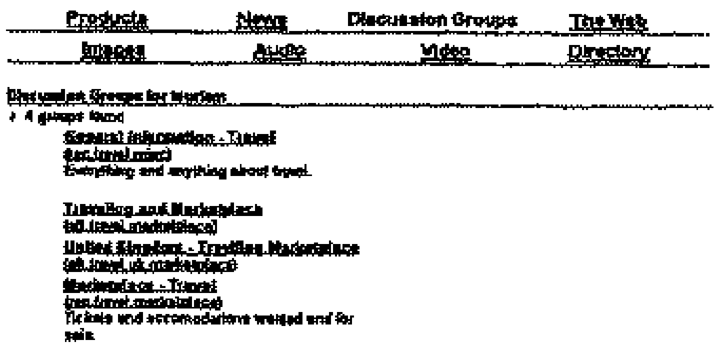
Books	(215 items)
Movies	(2 items)
Music	(4 items)

الشكل (6-39): نتائج البحث في AltaVista ضمن قسم المنتجات.

ي. مجموعات الحوار

يبين الشكل (6–40) نتائج البحث ضمن مجموعات الحوار عن tourism.

ننبه إلى ملاحظة أخيرة هي أن الأقسام الثمانية المشروحة سابقاً تنتج عند استخدام البحث العادي، بينما لا ينتج عن البحث المتقدم إلا قسم وحيد هو صفحات الوب.



الشكل (6–40): نتائج البحث في AltaVista ضمن قسم مجموعات الحوار.

6. لا يسمح بالتغذية الراجعة.

7. قاعدة البيانات

يمتلك AltaVista قاعدة بيانات ضخمة جداً، تصل إلى 150 مليون صفحة، ولكن رغم هذه الضخامة فإننا نجد أن نتائج البحث التي يعيدها مرضية وتبقى نسبة صفحات الضجيج مقبولة.

8. يتمتع AltaVista بسرعة جيدة، فلا نعاني من طول الانتظار من أجل الحصول على النتائج، وذلك رغم ضخامة قاعدة بياناته.

9. يؤمّن AltaVista مجموعة من Meta Words تسمح بالبحث عن معلومات غير نصية. يمكن الاستفادة من هذه الـ Meta Words في البحث، ضمن مؤشرات HTML الموجودة في الصفحات، عن معلومات مخصصة، مثل البحث في عنوان صفحة، أو في الارتباطات التشعبية ضمن الصفحة، أو URL أو... الجدول (6–2) التالي يعرض الـ Meta Words التي يدعمها AltaVista.

Meta Word	الشرح
anchor:*text*	البحث عن صفحات تحوي النص text ضمنها كارتباط تشعبي
applet:*class*	البحث عن الصفحات التي تستخدم البريمج (Java applet) *class*.
domain:*name*	يجد الصفحات ضمن مجال معين، مثلاً: domain:uk يعني إيجاد الصفحات من المملكة المتحدة فقط
host:*name*	إيجاد الصفحات على حاسوب معين، مثلاً: host:Ugarit.Shopping.com يعني البحث عن الصفحات التي على الحاسوب Ugarit في Shopping.com
image:*filename*	إيجاد الصفحات التي تحتوي الصورة *filename*
link:*URLtext*	إيجاد الصفحات التي تحتوي ارتباطاً تشعبياً إلى العنوان *URLtext*
text:*text*	إيجاد الصفحات التي تحتوي النص *text* في أي مكان في الصفحة عدا مؤثرات HTML الخاصة بالصور أو الارتباطات التشعبية أو URL
title:*text*	إيجاد الصفحات التي تحتوي النص *text* ضمن عنوان الصفحة (ضمن مؤشر HTML: <title></title>)
url:*text*	إيجاد الصفحات التي تحوي النص *text* ضمن الـ URL الخاص بها

جدول (6-2): مجموعة الـ **Meta Words** التي يدعمها **AltaVista**.

ج. البحث حسب الموصفات

يدعم AltaVista البحث حسب الموصفات ضمن مجموعة من أدواته مثل الصفحات الصفر، وأدوات التسوق التي يوفرها من أجل البحث عن منتج معين.

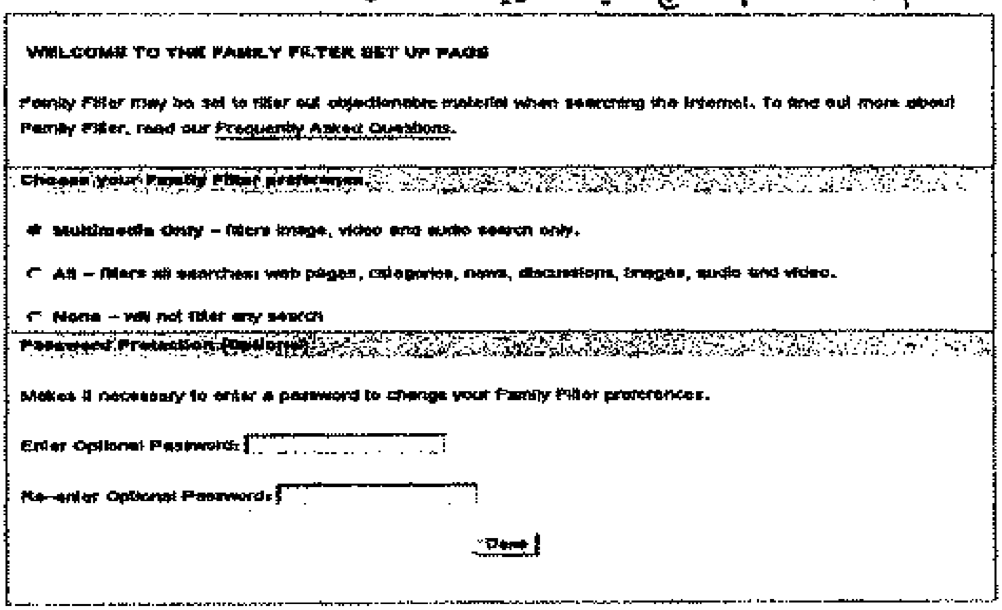
6-3-3 خدمات المعلومات الأخرى

1. Family Filter: هي ميزة هامة ومفيدة يقدمها AltaVista من أجل فلترة الصفحات ذات المواضيع غير المرغوبة من قبل المستخدم (الشكل 6-41).

المواضيع غير المرغوبة مصنفة في 5 فئات هي:

- المخدرات والكحول والتبغ.
- المقامرة.
- الكلام البذيء.
- الإباحة الجنسية.
- العنف.

يمكن للمستخدم أن يحدد قسم النتائج الذي ستجري عليه الفلترة.



الشكل (6-41): صفحة Family Filter

2. BABELFISH: أداة ترجمة من لغة إلى أخرى تسمح للمستخدم بأن يترجم نصوصه الخاصة إلى لغات أخرى، كما يمكنه ترجمة صفحات الويب بأن يحدد URL الخاصة بهذه الصفحة. يجب الانتباه إلى أن هذه الأداة ليست موثوقة جداً، كما أن الجمل المعقدة تسبب ترجمات غير صحيحة، لكنها تظل أداة مفيدة، فيمكن للمستخدم مثلاً أن يتبادل الرسائل مع أشخاص لا يعرف لغتهم، بأن يترجم رسائلهم باستخدام BABELFISH شريطة أن يستخدموا لغة سهلة وغير معقدة قواعدياً. اللغة العربية غير مدعومة من قبل هذه الأداة.

6-3-4 إضافة موقع إلى محرك البحث AltaVista

تمكن إضافة موقع إلى AltaVista من خلال الارتباط التشعبي على الصفحة الرئيسية، ضمن الجزء بعنوان Add a URL. يطلب AltaVista من المستخدم أن يعطيه URL موقعه، ثم يقوم عنكبوت AltaVista بزيارة الموقع، وفهرسة الصفحات التي ضمنه.

لا يحتاج AltaVista لأكثر من يوم واحد كي يضيف الموقع إلى فهرسه الخاص، لكنه قد يحتاج إلى ثلاثة أسابيع كي يضيف الموقع إلى جميع المواقع المرايا المشابهة له (Mirror sites).

إذا أراد المستخدم تحديد مجموعة كلمات مفتاحية تعبر عن مضمون الصفحة، أو إضافة شرح عن صفحاته (يظهر هذا الشرح ضمن صفحة نتائج البحث جانب عنوان صفحاته)، فيمكنه القيام بذلك باستخدام مؤشرات HTML الخاصة، والتي رأيناها عند الحديث عن HotBot. إذا لم يحدد المستخدم شرحاً عن صفحته، فإن AltaVista يعتبر الأسطر الأولى من الصفحة بمثابة شرح عن هذه الصفحة.

يمكن أن تمنع AltaVista من فهرسة مجموعة من الصفحات باستخدام مؤشرات HTML الخاصة، أو باستخدام الملف Robots.txt، أي بنفس الطريقة التي نتبعها مع HotBot. يدعم AltaVista أربعة أنواع من مؤشرات HTML الخاصة هي:

NOINDEX: يمنع فهرسة الصفحة.

NOFLLOW: يمنع عنكبوت AltaVista من ملاحقة الارتباطات التشعبية الموجودة ضمن الصفحة ومن فهرسة الصفحات التي على الطرف الآخر من الارتباط التشعبي.

NOIMAGEINDEX: يمنع فهرسة الصور ضمن الصفحة، ولكن يسمح بفهرسة النص.

NOIMAGECLICK: يمنع استخدام الارتباطات التشعبية إلى الصور مباشرة.

6–4 محرك البحث InfoSeek

infoseek

WWW.InfoSeek.com

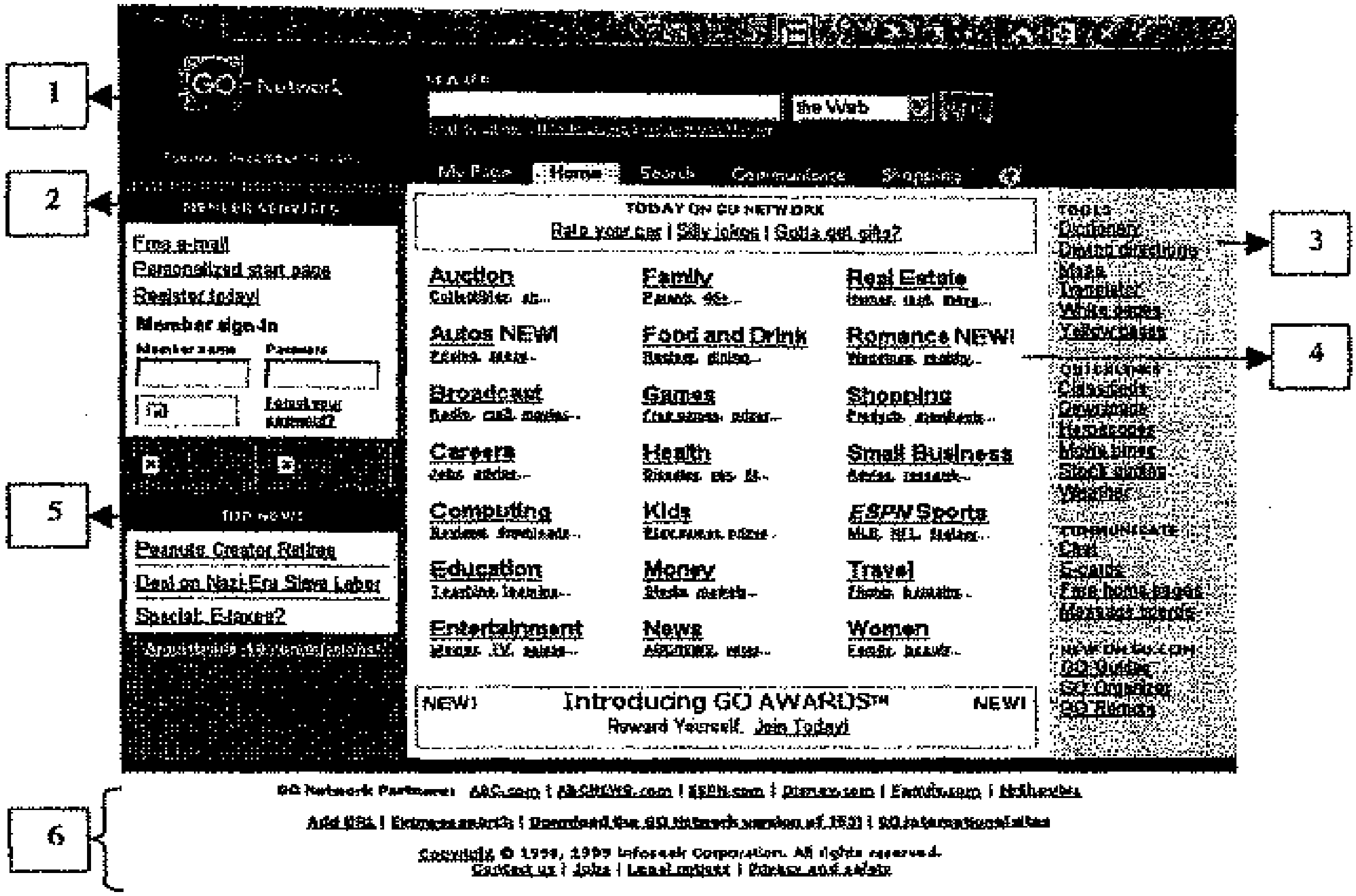
يعتبر InfoSeek من المحركات القوية والمفيدة في البحث عن المعلومات في الإنترنت، فهو يمتلك فهرساً لـ 75 مليون موقع، كما أنه يمتلك مجلداً يدوياً تُصنّف فيه المواقع حسب مواضيعها. يقوم أشخاص مختصون ببناء مجلد InfoSeek، حيث يفحصون المواقع التي يطلب المستخدمون تسجيلها في مجلد InfoSeek، ليختاروا منها المواقع الجيدة، وليضيفوها إلى المجلد.

تُصنّف المواقع ضمن مجلد InfoSeek إلى الأقسام الأربعة التالية:

- *المواقع الأفضل* Best Sites: هي المواقع التي لا بد أن تكون مهمة للمستخدم. ويجب أن تكون غزيرة المعلومات (أو مسلية)، وسهلة الاستخدام والتصفح، وعالية التنظيم.
- *المواقع الجيدة جداً* Very good sites: هي مواقع غنية المضمون، وذات تصميم وتنظيم جيدين، كما أنها تتصف بسهولة الاستخدام.
- *المواقع الجيدة* Good sites: وتشمل باقية مواقع المجلد، وهي مواقع جيدة المضمون، وسهلة الاستخدام.
- بعض المواقع المتخصصة مثل ABCNEWS.com وESPN.com.

6–4–1 واجهة InfoSeek

يُظهر الشكل (6–42) صفحة InfoSeek على الويب، والتي يمكن الوصول إليها باستخدام العنوان: http:\\www.infoseek.com ؛ نجد فيما يلي الإمكانات المختلفة لـ InfoSeek:



الشكل (6–42): صفحة Infoseek

- الجزء [1]: يسمح بالقيام بعمليات بحث، ويمكن أن تحدد منطقة البحث باستخدام المنزلقة بجانب زر Find.
- الجزء [2]: يسمح للمستخدم بأن ينضم إلى مجتمع InfoSeek، ويؤمن مجموعة من الأدوات للأعضاء.
- الجزء [3]: مجموعة من أدوات InfoSeek.
- الجزء [4]: مجلد InfoSeek.
- الجزء [5]: أخبار اليوم.

- الجزء [6]: مجموعة من الأدوات لإضافة موقع إلى InfoSeek، وارتباطات إلى المشاركين في شبكة Go.

6-4-2 البحث عن المعلومات في InfoSeek

يتيح InfoSeek الأنواع الثلاثة من طرق البحث عن المعلومات: التصفح، البحث عن المحتوى، البحث عن الموصفات.

آ. التصفح

يشابه البحث بالتصفح في InfoSeek التصفح في باقي محركات البحث، فهو يملك مجلداً خاصاً تُصنَّف المواقع ضمنه حسب مواضيعها. ويمكن التصفح فيه باستخدام الارتباطات التشعبية ضمن الجزء [4].

ب. البحث عن المحتوى

يتيح InfoSeek إمكانية صياغة طلبات البحث من خلال الجزء [1] من صفحته الرئيسية. يُظهر الشكل (6-43) واجهة البحث هذه، وسنعرض مزاياها المختلفة فيما يلي:



الشكل (6-43): واجهة البحث ضمن الصفحة الرئيسية لـ InfoSeek

- الحقل الحواري هو مكان صياغة طلبات المستخدم.
- تحدد المنزلقة منطقة القيام بالبحث، فيمكن البحث في:
  - The Web: كامل الوب.

- Topics: مجلد مبني يدوياً يضمّ أفضل المواقع على الشبكة.
- News: الأخبار.
- Companies: المعلومات الخاصة بالشركات.
- Newsgroups: مجموعة مجموعات الأخبار.

نجد أسفل حقل الحوار ارتباطاً تشعبياً إلى البحث المتقدم، وصفحات مساعدة للقيام بالبحث، والأداة GOguardian التيسنتحدث عنها لاحقاً في هذا الفصل.

تقييم البحث بالمحتوى

1. البحث البسيط

يسمح InfoSeek بعمليات البحث البسيط، ومن أجل فحص إمكانات هذا النوع من البحث، قمنا بالتجارب التالية:

- التحسس للأحرف الصغيرة والكبيرة: بالبحث عن tourism و Tourism، نجد اختلافاً في عدد صفحات المادة كلّ مرة. مما يعني أنه يتحسس للأحرف الكبيرة والصغيرة. وهو يتحسس للأحرف على النحو الذي يتحسس فيه AltaVista لها.
- البحث عن جمل: بالبحث عن "tourism in Syria"، نجد أن الصفحات التي وجدها النظام تتحدث عن السياحة في سورية، وتحوي هذه الجملة. إذن يتيح InfoSeek إمكانية تجميع كلمات البحث في جمل، إذ يكفي وضع الجملة ضمن "" كي يقوم بعملية بحث عن الصفحات التي تحوي الجملة بكاملها.
- الكلمات المفتاحية اللازمة: يسمح InfoSeek للمستخدم بأن يحدد كلمة يريد لها أن تكون محتواة ضمن جميع الصفحات المطابقة. فمثلاً يؤدي الطلب: tourism +Syria إلى إيجاد الصفحات التي تتحدث عن tourism وتحتوي ضمنها كلمة Syria بالضرورة.

- الكلمات المفتاحية الممنوعة: يتيح InfoSeek إمكانية منع كلمة من أن تكون موجودة ضمن صفحة ما عن طريق استخدام "-". فمثلاً: بالبحث عن "tourism" والبحث عن "- tourism Syria"، نجد بفحص النتائج أن الصفحات التي تتحدث عن "Syria" قد اختفت في الطلب الثاني.
- لا يسمح باستخدام الرموز العامة (Wild Card).
- يرتب النتائج حسب درجة المطابقة المتناقصة، التي يحسبها باستخدام تابع مطابقة خاص.

لا يسمح InfoSeek باستخدام البحث المتقدم (حسب تعريفنا له) ويكتفي بإمكانيات منع أو لزوم الكلمات بدلاً عن العوامل المنطقية.

2. البحث المتقدم

يُقسم البحث المتقدم إلى مجموعة من الأقسام (يمكن الوصول إليه من الارتباط التشعبي Search options أسفل حقل الحوار)، ويمكن البحث ضمن أي قسم منها. بمعنى آخر، البحث المتقدم في InfoSeek ليس بحثاً متقدماً حسب تعريفنا، وإنما تخصيصٌ لمجالات البحث بهدف الحصول على نتائج أفضل. تظهر في الشكل (6-44) أقسام البحث المتقدم في InfoSeek:



الشكل (6-44): أقسام البحث المتقدم في InfoSeek

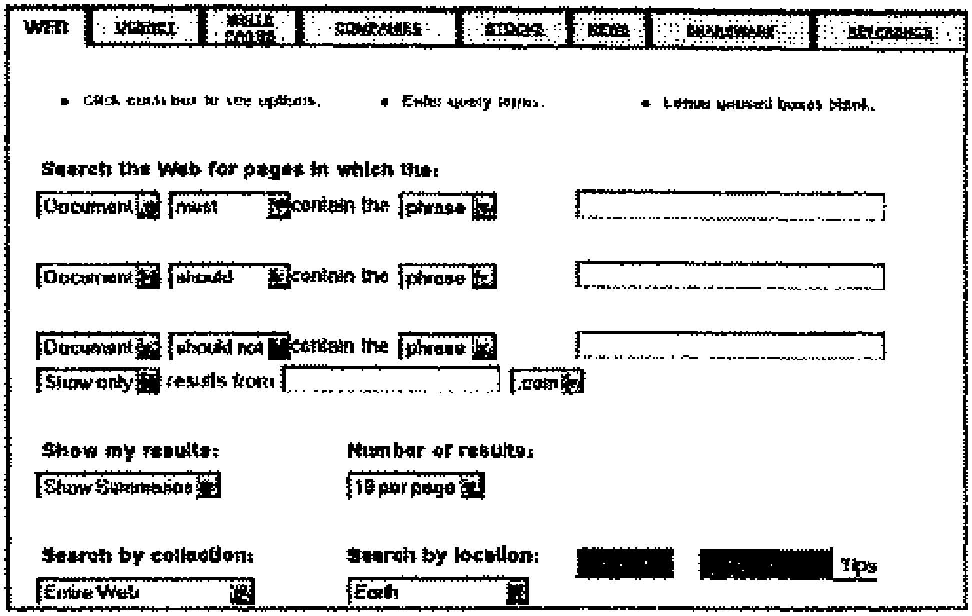
آ. الوب Web

نلاحظ أن طريقة صياغة الطلبات مختلفة عن المألوف، فالمستخدم يصيغ طلباته ضمن حقل الحوار الموجود في يمين وأعلى الشكل (6-45) كما يلي:

ضمن المنزلقة الأولى، يحدد المستخدم أين يريد أن يبحث InfoSeek عن الطلب: هل ضمن الصفحة، أم عنوانها، أم URL، أم في الارتباطات التشعبية ؟

ضمن المنزلقة الثانية، يحدد المستخدم إن كانت هذه الكلمات يجب أن توجد أو لا توجد ضمن الصفحات النتائج (كلمات لازمة أو ممنوعة). ضمن المنزلقة الثالثة، يحدد المستخدم نوع الطلب الذي يصيغه: هل هو جملة، أم اسم علم، أم كلمات ؟ يمكن للمستخدم أن يصوغ ثلاثة طلبات معاً من أجل نفس البحث.

بعد ذلك، يمكن للمستخدم أن يحدد عمليات فلترة على النتائج، فبإمكانه، باستخدام المنزلقة التي يظهر فيها Show only، أن يحدد موقعاً معيناً للبحث ضمنه فقط، أو لاستثنائه من عملية البحث، ويجري تحديد الموقع ضمن حقل الحوار الذي يلي.



الشكل (6–45): البحث المتقدم في **InfoSeek** ضمن قسم الوب

يمكن التحكم بعملية إظهار النتائج بأن يحدد المستخدم من المنزلقة ذات العنوان Show my results رغبته (أو عدم رغبته) في الحصول على شرحٍ عن كل نتيجة. كما يمكنه باستخدام المنزلقة التالية أن يحدد عدد النتائج التي يريدها على كل صفحة من صفحات النتائج. أخيراً تسمح المنزلقتان في آخر الشكل (6–45) للمستخدم بأن يحدد مجال البحث ضمن موضوع معين (ثقافة، صحة، أخبار...)،

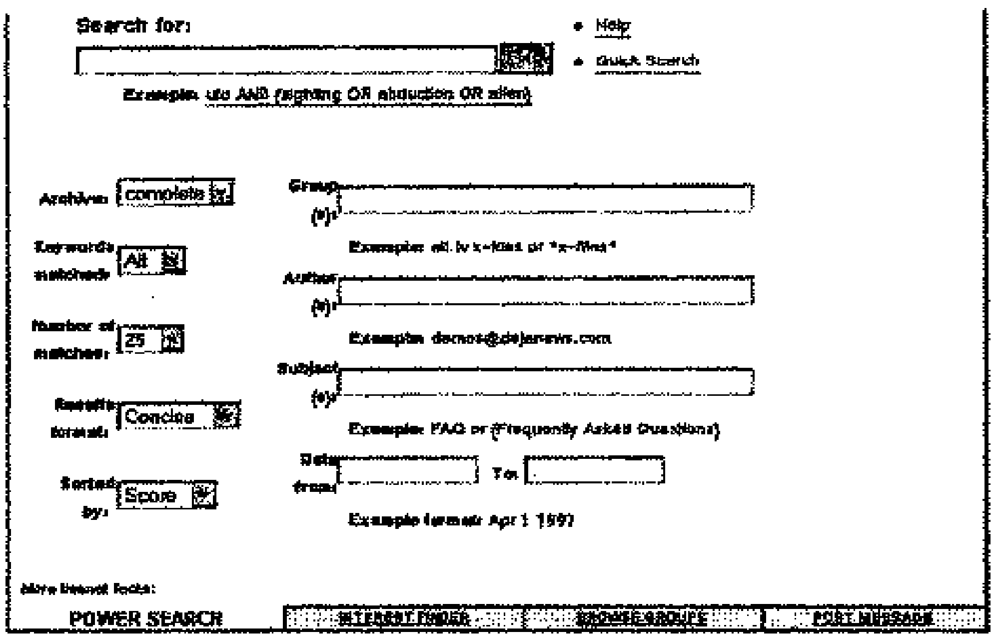
كما يمكن أن يحدد البحث ضمن مكان جغرافي معين (آسيا، إفريقية، أو دولة من الدول...)، أو أن يحدد مجالاً من مجالات الأسماء على الإنترنت مثل (com, .edu, .gov,...).

ب. مجموعة المجموعات الأخبارية Usenet

يمكن البحث فيها باستخدام العوامل المنطقية AND, OR، ويمكن القيام بمجموعة كبيرة من عمليات الفلترة حسب الأرشيف والمجموعة الأخبارية والكاتب والموضوع، إضافةً إلى الفلترة حسب التاريخ (مجال زمني معين) كما يُظهر الشكل (6–46). توجد كذلك مجموعة من الارتباطات التشعبية في أسفل الصفحة تسمح بالقيام بعمليات بحث أكثر تخصصاً ضمن مجموعة المجموعات الأخبارية.

ج. الصفحات البيض White Pages

تسمح عمليات البحث ضمن هذا القسم بالبحث عن الأشخاص، وهي طريقة خاصة من البحث باستخدام الموصفات، وسنتحدث عنها ضمن البحث عن الموصفات.



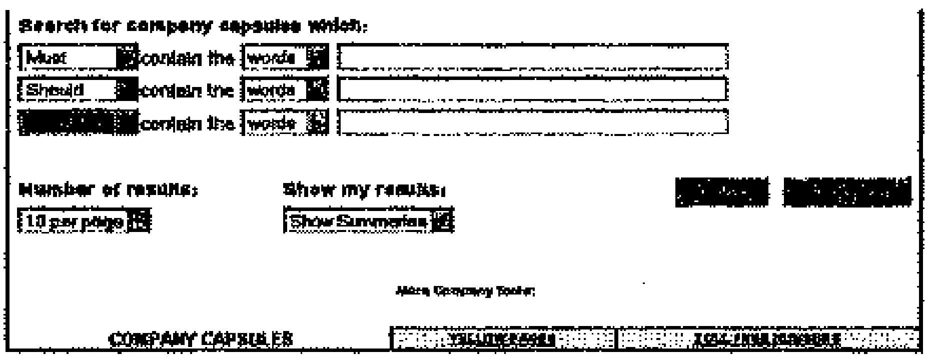
الشكل (6–46): البحث المتقدم في Infoseek ضمن قسم مجموعة مجموعات الأخبار

د. *الشركات Companies*

تسمح بصياغة طلبات باستخدام العوامل البوليانية من أجل استحضار صفحات عن الشركات تلبي على طلب المستخدم. نجد في أسفل الشكل (6–47) مجموعة من الارتباطات للقيام بعمليات بحث أخرى.

هـ. *البضائع Stocks*

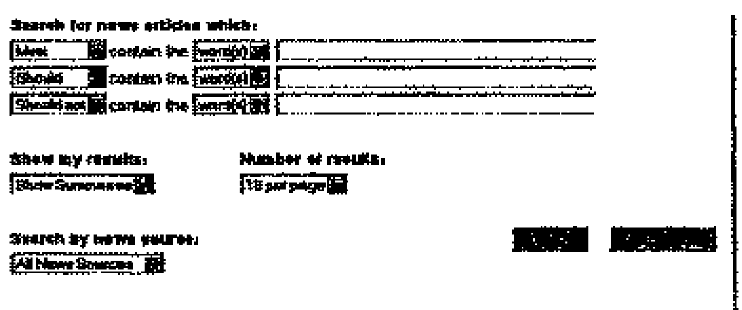
أداة بحث بالموصفات عن البضائع.



الشكل (6–47): البحث المتقدم في InfoSeek ضمن قسم الشركات

و. *الأخبار News*

يمكن ملاحظة أن البحث في هذا القسم مشابه جداً للبحث في قسم الوب (انظر الشكل 6–48)، حيث يحدد المستخدم طلب البحث ضمن مربعات الحوار على النحو نفسه الذي قام به في قسم الوب، ويتحكم بالإظهار من حيث عدد النتائج، وعرض الشرح لكل موقع، كما يمكن تحديد مكان القيام بالبحث باستخدام المنزلقة الأخيرة.



الشكل (6–48): البحث المتقدم في Infoseek ضمن قسم الأخبار

ز. البرمجيات التشاركية Shareware

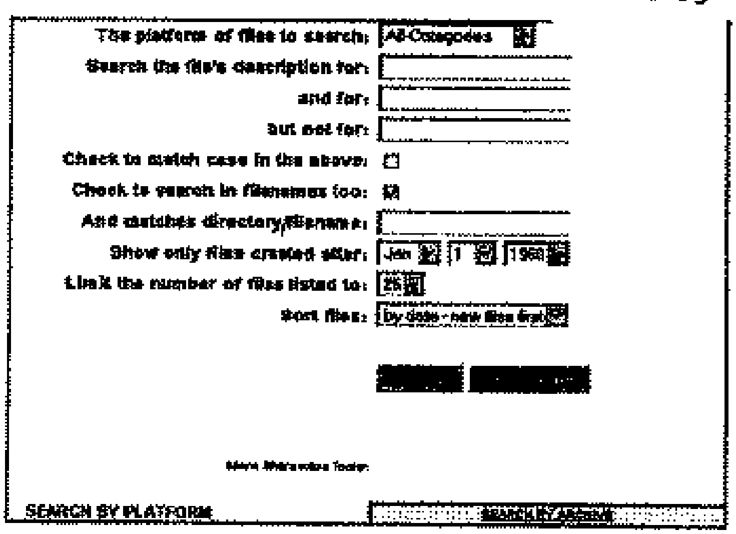
تسمح هذه الأداة بالبحث عن البرمجيات التشاركية، ويقوم المستخدم بإدخال الطلب ضمن حقول الحوار أعلى الصفحة (تحت المنزلقة مباشرة) في الشكل (6-49)، حيث يبحث InfoSeek بين الشروح التي يحويها عن البرمجيات التشاركية ليجد البرمجيات المطابقة. يمكن القيام بمجموعة من عمليات الفلترة مثل الفلترة حسب بيئة عمل معينة، وحسب التاريخ، ...

ي. Reference

تسمح باستخدام أدوات InfoSeek بالبحث عن خرائط الطرق، والترجمة والمرادفات، وأسماء المجالات، والأسماء المختصرة، وهي أدوات بحث حسب الموصفات.

3. يسمح بالفلترة حسب الموضوع والمنطقة الجغرافية ومجالات الأسماء على الإنترنت، وفي بعض أدوات البحث المتقدم يسمح بالفلترة حسب التاريخ. كما يسمح أيضاً بالفلترة حسب مواضيع الصفحات باستخدام GOguardian الذي يمنع الصفحات التي تحوي مواضيعَ سيئةً للأطفال (سنتحدث عنه لاحقاً).

4. يؤمن تصنيفاً جيداً للمواضيع ضمن فئات بفضل مجلده المبني يدوياً، ونجد الفئات الرئيسية لمجلده ضمن الجزء [4] من الصفحة الرئيسية.



الشكل (6-49): البحث المتقدم في InfoSeek ضمن قسم البرمجيات التشاركية

5. إظهار النتائج

يُظهر الشكل (6–50) صفحة النتائج التي يعيدها InfoSeek، وهي تحتوي الأجزاء التالية:

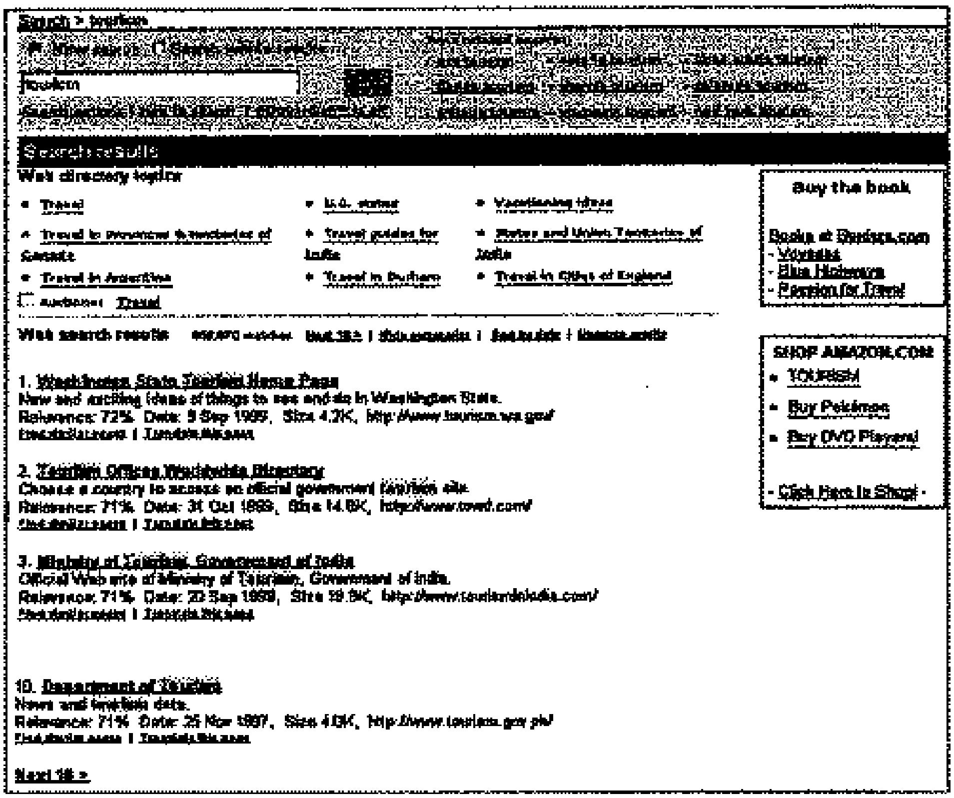
أ. طلبات مقترحة من InfoSeek (Try a related search): يشتق InfoSeek طلبات بحث من طلب المستخدم الأصلي، ويقترح هذه الطلبات على المستخدم الذي قد يجد في بعضها شيئاً مفيداً. وقد رأينا سابقاً أن HotBot وAltaVista يقومان بالعمل نفسه.

ب. نتائج البحث ضمن مجلد InfoSeek (Web directory topics): وهي الفئات التي وجدها مطابقة لطلب المستخدم ضمن مجلده الخاص. بجانب هذه الفئات، يعرض InfoSeek مجموعة من الكتب التي تطابق طلب المستخدم أيضاً.

ج. نتائج البحث في صفحات الوب (Web search results): وهي الصفحات المطابقة لطلب المستخدم والتي وجدها InfoSeek على الوب. نلاحظ ضمن هذه النتائج ما يلي:

- عدد الصفحات المطابقة التي وجدها InfoSeek، وهو في مثالنا 555870 صفحة.
- عنوان الصفحة النتيجة كارتباط تشعبي إلى الصفحة، ونلاحظ أن كلمة tourism (وهي طلب المستخدم) تظهر بخلفيةٍ صفراءَ فاتحةٍ (لا تظهر على الصورة غير الملونة).
- تحت العنوان يوجد شرح بسيط عن هذه الصفحة. وضمن هذا الشرح نجد أن كلمة tourism قد ظهرت داكنة وبخلفية صفراء فاتحة (تُلاحظ في النتيجة الثانية).
- نسبة مئوية هي نسبة مطابقة الصفحة لطلب المستخدم. يقوم InfoSeek بحساب هذه النسب اعتماداً على خوارزميات المطابقة التي يستخدمها. ولا يُعطي InfoSeek أية معلومة عن خوارزمياته، ولكن حسب معرفتنا بأنظمة استرجاع المعلومات وخوارزمياتها، يمكن أن نتوقع أن نموذج المطابقة المستخدم هو النموذج الشعاعي. من أجل النتيجة الأولى، نجد أن نسبة المطابقة هي 71٪.
- بجانب نسبة المطابقة، نجد تاريخ آخر تعديل للصفحة، وحجمها، وURL.

* يُقدم InfoSeek ميزة خاصة تكمن في الارتباط التشعبي Find similar pages. يقودنا هذا الارتباط إلى مجموعة من الصفحات التي تُطابق الصفحة الحالية. إن وجود هذه الميزة يعزز اعتقادنا بأن نموذج المطابقة المستخدم هو النموذج الشعاعي.



الشكل (6–50): صفحة النتائج في InfoSeek

د. يعرض InfoSeek 10 نتائج في كل صفحة نتائج، وهو رقم يمكن تغييره من قبل المستخدم. وفي أسفل صفحة النتائج، نجد ارتباطاً تشعبياً إلى النتائج العشر التالية (Next 10).

6. يسمح بالتغذية الراجعة، حيث نجد في أعلى صفحة النتائج، كما يوضّح الشكل (6–51)، خياراً من أجل البحث ضمن نتائج آخر عملية بحث، أو تجديد البحث دون أي اعتبار للنتائج السابقة.



الشكل (6-51): التغذية الراجعة في InfoSeek.

7. يمتلك InfoSeek قاعدة بيانات ضخمة.

8. يتمتع InfoSeek بسرعة جيدة، فلا نعاني من طول الانتظار من أجل الحصول على النتائج، رغم ضخامة قاعدة بياناته.

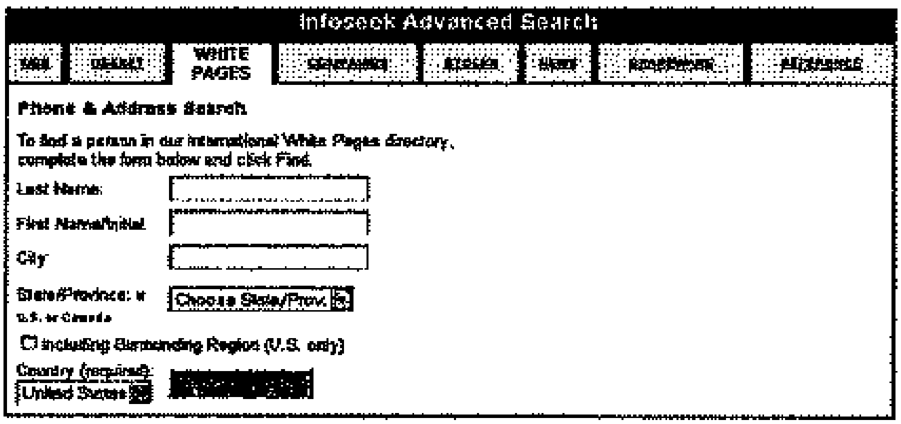
9. يؤمن InfoSeek مجموعة من الطرق للبحث بين URL الصفحات، أو العناوين، أو ضمن الارتباطات التشعبية. وفيما يلي طرق البحث المختلفة:

- البحث ضمن العناوين: الطلب title:tourism، يبحث عن الصفحات التي تحوي tourism ضمن عناوينها.
- البحث ضمن URL: الطلب url:tourism، يبحث عن الصفحات التي تحتوي كلمة tourism ضمن URL الخاصة بها.
- البحث ضمن موقع: الطلب site:microsoft.com Bill Gates، يبحث عن الصفحات التي تحوي Bill Gates ضمن الموقع microsoft.com فقط.
- البحث عن الصفحات المرتبطة بموقع معين: الطلب link:microsoft.com يبحث عن الصفحات التي ترتبط بالموقع microsoft.com.
- البحث ضمن نتائج بحث سابق: الطلب tourism +Syria | Aleppo، يبحث عن الصفحات التي تتحدث عن السياحة السورية في حلب فقط.
- استخدام اسمَي علم في الطلب: يتم الفصل بينهما باستخدام فاصلة , (،) كما في المثال: George Bush, Texas. حيث George Bush اسم علم، وTexas اسم علم آخر.

تُذكِّر المستخدم أن البحث المتقدم يُقدم معظم هذه الإمكانات من خلال قسم الويب.

ج. البحث حسب الموصفات

يدعم InfoSeek البحث حسب الموصفات ضمن مجموعة من أدواته مثل الصفحات البيض التي تسمح بالبحث عن الأشخاص، فالمستخدم يُدخل اسم الشخص أو كنيته ضمن الحقول المخصصة (والتي توافق حقولاً في قاعدة البيانات)، ثم يقوم InfoSeek بالبحث ضمن قاعدة بياناته عن القيم الموافقة، ويعرض النتائج. يبين الشكل (6–52) الصفحات البيض ضمن البحث المتقدم في InfoSeek.



الشكل (6–52): الصفحات البيض في InfoSeek

6–4–3 خدمات المعلومات الأخرى

1. Dictionary: وهو أداة تسمح باستخدام قاموس Webster من أجل التحقق من تهجئة الكلمات أو الحصول على شرح لها. كما يمكن استخدام مكنز Webster من أجل الحصول على مرادفات لكلمة معينة. يبين الشكل (6–53) واجهة هذه الأداة.

2. GOguardian: وهو أداة تفيد في فلترة الصفحات التي تحتوي أموراً يجب على الأطفال عدم الاطلاع عليها، كتلك التي تتعلق بالعنف والجنس والمخدرات. يقوم GOguardian بمراقبة نتائج

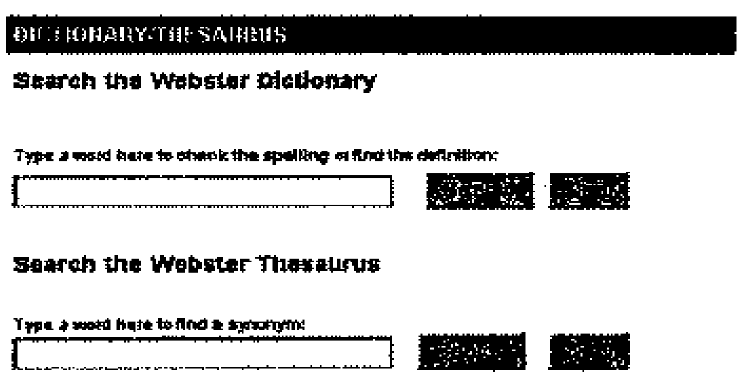
البحث المعادة، ويتخلص من الصفحات التي تحوي أموراً ممنوعة، مما يضمن حماية الأطفال من تأثيرها أثناء قيامهم بعمليات بحث، كما يوفر إمكانية إضافة كلمة سر، مما يُعطي ضمانة أكيدة بعدم قدرة أحد على تعديل الإعدادات التي ضبطها المستخدم.

6–4–4 إضافة موقع إلى محرك البحث InfoSeek

تمكن إضافة موقع إلى InfoSeek من خلال الارتباط التشعبي Add a URL في الصفحة الرئيسية ضمن الجزء 6. يطلب InfoSeek من المستخدم أن يعطيه URL موقعه، ثم يقوم عنكبوت InfoSeek بزيارة الموقع، وفهرسة الصفحات التي ضمنه.

يحتاج InfoSeek إلى حوالي 6 أسابيع حتى يضيف الموقع إلى فهرسه، ويمكن منعه من فهرسة صفحات معينة باستخدام الملف Robots.txt (كما في محركات البحث السابقة). ننوه إلى عدم توفر إمكانية منع عملية فهرسة صفحات معينة باستخدام مؤشرات HTML خاصة (كما هو الحال مع محركات البحث الأخرى).

يمكن للمستخدم أن يحدد مجموعة كلمات مفتاحية تعبر عن مضمون الصفحة، أو أن يضيف شرحاً عن صفحاته (يظهر هذا الشرح ضمن صفحة نتائج البحث جانب عنوان صفحاته)، وذلك باستخدام مؤشرات HTML الخاصة، وهي keywords من أجل إضافة الكلمات المفتاحية، وdescription من أجل الشرح (رأيناها سابقاً عند الحديث عن HotBot). يمكن إضافة شرح بطول 200 حرف فقط، أما الكلمات المفتاحية، فيمكن إضافة كلمات حتى طول 1000 حرف.



الشكل (6–53): واجهة الأداة Dictionary في InfoSeek

5-8 محرك البحث Lycos

WWW.Lycos.com

من المحركات الضخمة على الوب، فهو يمتلك فهرساً لـ 100 مليون صفحة. يتميز بقدراته الكبيرة في مجال البحث المقارب. يُعتبر محرك Lycos أحد أهمّ الشركاء في شبكة Lycos (Lycos Network)، وهي أكثر الشبكات زواراً في العالم (بين الشبكات المماثلة). تُقدّم هذه الشبكة إضافةً إلى عمليات البحث عن المعلومات مجموعة كبيرة من الخدمات مثل: chat, E-mail, online sopping, personalized new,... ننوه هنا إلى أن المحرك HotBot هو أحد أعضاء هذه الشبكة أيضاً.

يمتلك Lycos مجموعة من الصور الضخمة تُقدر بحوالي 80000 صورة، ويسمي هذه المجموعة Image Gallery. هذه الصور متعددة المواضيع: صور مشاهير، وصور أجهزة، وصور حيوانات. يسمح هذا المعرض للمستخدم بالتصفح ضمنه، لكن يمنعه من الاستخدام التجاري لهذه الصور. يمكن للمستخدم أيضاً القيام بعمليات بحث بين هذه الصور، فيدخل كلمات مفتاحية للموضوع الذي يريد مشاهدة صور له، عندئذٍ يبحث Lycos ضمن معرضه عن صور مطابقة. يمكن أيضاً التوسع في مجال البحث ليشمل الوب كاملة إذا لم يجد المستخدم طلبه ضمن المعرض.

6-5-1 واجهة Lycos

يُظهر الشكل (6-54) صفحة Lycos على الوب، والتي يمكن الوصول إليها باستخدام العنوان: http:\\www.lycos.com، نجد فيما يلي الإمكانات المختلفة لـ Lycos:

الجزء [1]: يسمح للمستخدم بالقيام بعمليات بحث.

الجزء [2]: آخر أخبار Lycos.

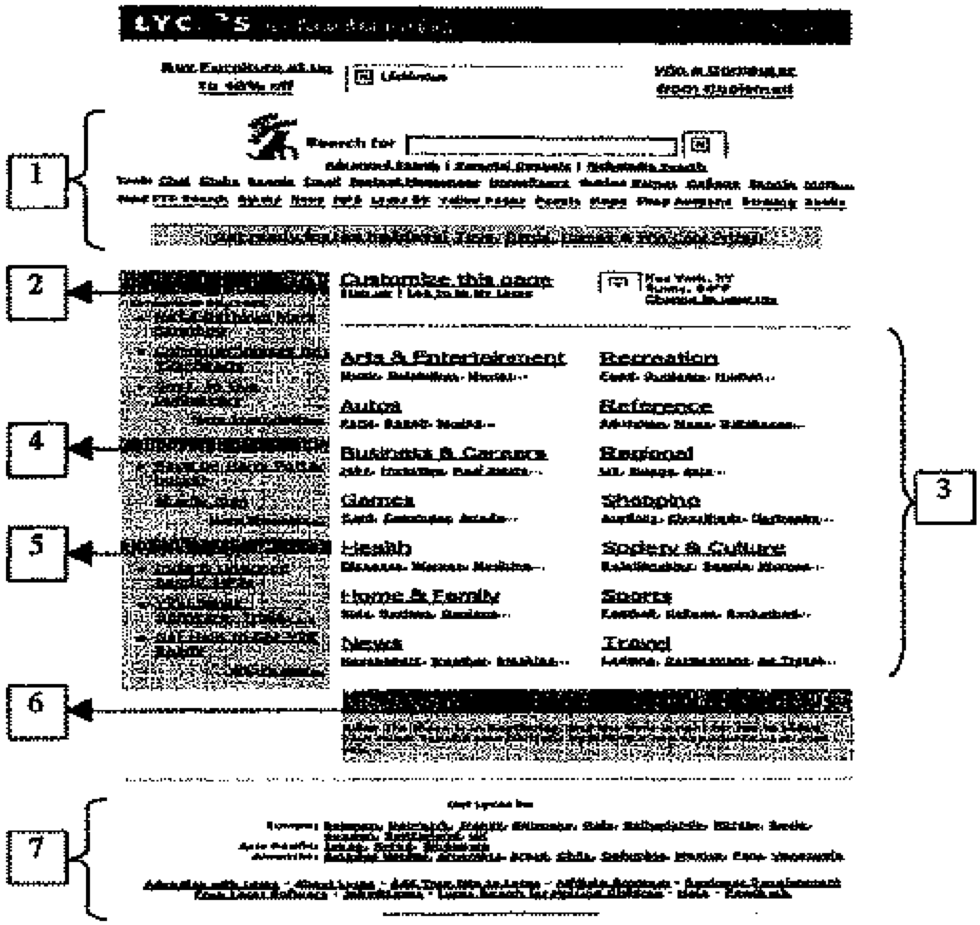
الجزء [3]: مجلد Lycos.

الجزء 4: التسوق باستخدام Lycos.

الجزء 5: مواضيع ومقالات يقدمها Lycos.

الجزء 6: توقعات وأبراج فلكية.

الجزء 7: مواقع لـ Lycos ضمن مجموعة من الدول.



الشكل (6-53): صفحة Lycos

6-5-2 البحث عن المعلومات في Lycos

يتيح Lycos الأنواع الثلاثة من طرق البحث عن المعلومات: التصفح، البحث عن المحتوى، البحث عن الموصفات.

آ. التصفح

يشابه البحث بالتصفح في Lycos التصفح في باقي محركات البحث، فهو يمتلك مجلداً خاصاً تُصنّف المواقع ضمنه حسب مواضيعها. ويمكن التصفح فيه باستخدام الارتباطات التشعبية ضمن الجزء [3].

ب. البحث عن المحتوى

يتيح Lycos إمكانية صياغة طلبات البحث من خلال الجزء [1] من صفحته الرئيسية.

تقييم البحث بالمحتوى

1. البحث البسيط

يسمح Lycos بعمليات البحث البسيط، ومن أجل فحص إمكانات هذا النوع من البحث، قمنا بالتجارب التالية:

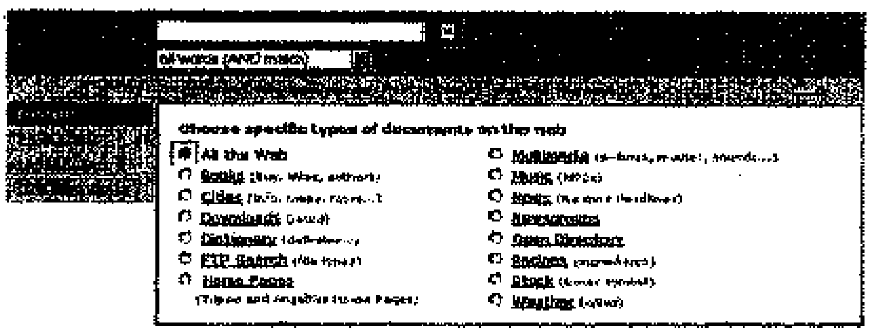
- التحسس للأحرف الصغيرة والكبيرة: بالبحث عن tourism و Tourism، نجد اختلافاً في عدد صفحات المعادة كلّ مرة. مما يعني أنه يتحسس للأحرف الكبيرة والصغيرة. يتحسس للأحرف على النحو الذي يتحسس فيه AltaVista لها.
- البحث عن جمل: بالبحث عن "tourism in Syria"، نجد أن الصفحات التي وجدها النظام تتحدث عن السياحة في سورية، وتحوي هذه الجملة. إذن يتيح Lycos إمكانية تجميع كلمات

البحث في جمل، إذ يكفي وضع الجملة ضمن "" كي يقوم بعملية بحث عن الصفحات التي تحوي الجملة بكاملها.

- الكلمات المفتاحية اللازمة: يسمح Lycos للمستخدم بأن يحدد كلمة يريد لها أن تكون محتواة ضمن جميع الصفحات المطابقة. فمثلاً يؤدي الطلب: tourism +Syria إلى استحضار الصفحات التي تتحدث عن tourism وتحتوي ضمنها بالضرورة كلمة Syria.
- الكلمات المفتاحية الممنوعة: يتيح Lycos إمكانية منع كلمة من أن تكون موجودة ضمن صفحة ما عن طريق استخدام "-". فمثلاً: بالبحث عن "tourism" والبحث عن "tourism -Syria"، نجد بفحص النتائج أن الصفحات التي تتحدث عن "Syria" قد اختفت من نتائج الطلب الثاني.
- لا يسمح باستخدام الرموز العامة (Wild Card).

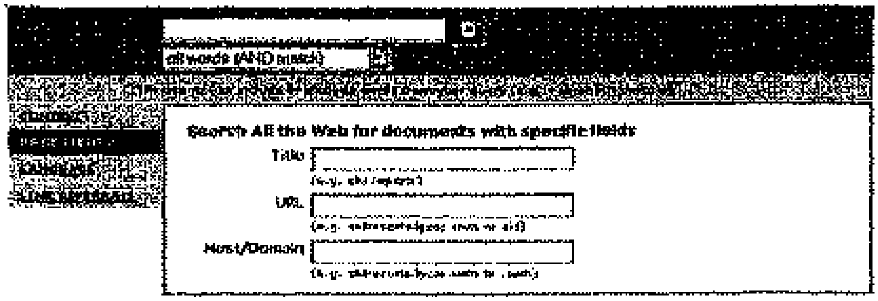
2. البحث المتقدم

إن صفحة البحث المتقدم (Advanced search) في Lycos تحوي إمكانات فلترة حسب الموضوع، كما نرى في الشكل (6-54). كما يمكن للمستخدم أن يستخدم العوامل البوليانية في صياغة طلباته.



الشكل (6-54): البحث المتقدم في Lycos حسب الموضوع

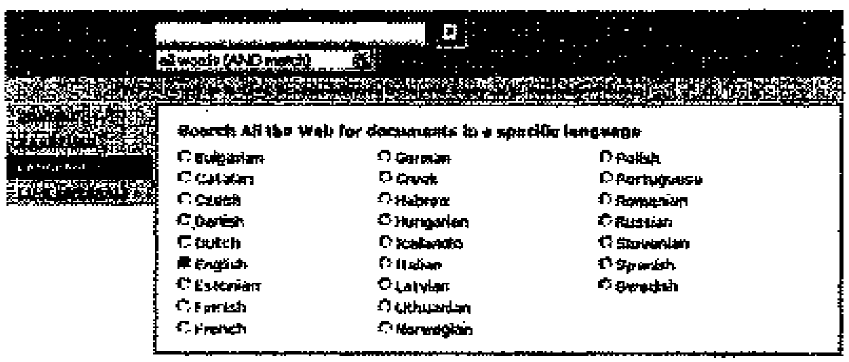
يمكن البحث حسب عنوان الصفحة، وURL وDomain، كما نرى في الشكل (6-55).



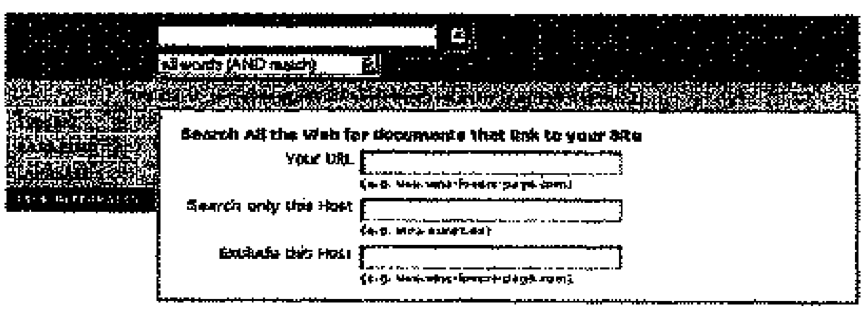
الشكل (6-55): البحث المتقدم في Lycos حسب حقول معينة في الصفحة

يمكن البحث عن صفحات مكتوبة بلغة معينة (فلترة حسب اللغـة)، كما ترى في الشـكل (6-56)، ونلاحظ أنه لا يدعم اللغة العربية.

كما يمكن البحث عن الصفحات التي ترتبط بموقع ما (موقع المستخدم مثلاً)، مع إمكانية فلترة النتائج حسب الموقع (أو منع إظهار النتائج الواردة من موقع ما)، كما ترى ذلك في الشكل (6-57).



الشكل (6-56): البحث المتقدم في Lycos حسب لغة الصفحة



الشكل (6-57): البحث المتقدم في Lycos حسب الموقع

3. البحث البولياني

يُتيح Lycos القيام بالبحث البولياني باستخدام العوامل البوليانية: AND, OR, NOT. فهو يسمح بصياغة طلبات، مثل:

(Tourism AND Syria) AND (NOT Aleppo)

من أجل البحث عن الصفحات التي تتحدث عن السياحة في سورية عدا تلك التي تتحدث عن حلب.

4. البحث المقارب

يدعم Lycos البحث المقارب دعماً كبيراً، ولعله أكثر محركات البحث الأكثر دعماً لهذا النوع من البحث، فهو يسمح باستخدام مجموعة من العوامل هي:

- ADJ: تسمح بالبحث عن صفحات تكون فيها كلمات الطلب متجاورةً دون مراعاةٍ للترتيب، فمثلاً الطلب: car ADJ race، يعيد صفحات تحوي car race أو race car.
- NEAR: تسمح بالبحث عن الصفحات التي يكون فيها عدد الكلمات المحصورة بين كلمتي الطلب أقل من 25 كلمة.
- FAR: تعمل عكس عمل NEAR، فهي تسمح بالبحث عن الصفحات التي يكون فيها عدد الكلمات المحصورة بين كلمتي الطلب أكبر من 25 كلمة.
- BEFORE: تعمل على نحوٍ مشابه جداً لعمل AND، فهي تسمح بالبحث عن الصفحات التي تحتوي كلمتي الطلب، لكنها تفرض ترتيباً لورود هاتين الكلمتين.
- مثلاً يعيد الطلب: race BEFORE car الصفحات التي تحتوي race car، لكنه لا يعيد الصفحات التي تحتوي car race.

نلاحظ إذن أن BEFORE هي العامل الوحيد الذي يفرض ترتيباً على ورود كلمات الطلب. يمكن استخدام العوامل السابقة لكن مع إضافة ترتيب على ورود كلمات الطلب باستخدام العوامل: ONEAR, OFAR, OADJ (الحرف O من Order أي ترتيب).

ننوه أخيراً إلى أن العاملَين NEAR, FAR يستخدمان عدد الكلمات 25 كعتبة قبول أو رفض للصفحة. يمكن أن نعدّل هذا العدد باستخدام "/". مثلاً يعيد الطلب Tourism NEAR/3 Syria، جميع الصفحات التي لا يتجاوز فيها عدد الكلمات بين Tourism وSyria ثلاث كلمات. يمكن أيضاً استخدام "/" مع ADJ, ONEAR, OFAR, OADJ بالطريقة نفسها.

5. يسمح Lycos بالفلترة حسب الموضوع فقط ضمن صفحة البحث المتقدم Advanced search، كما رأينا سابقاً في الشكل (6–54).

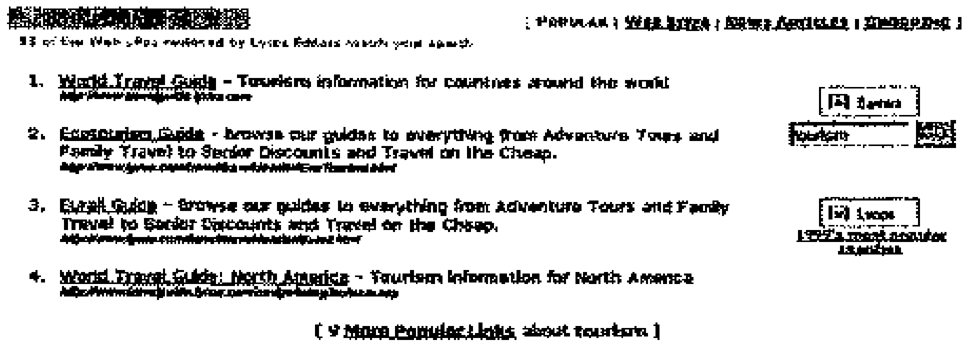
6. يؤمن تصنيفاً جيداً للمواضيع ضمن فئات بفضل مجلده المبني يدوياً، ونجد الفئات الرئيسية لمجلده ضمن الجزء 5 من الصفحة الرئيسية.

7. إظهار النتائج

يُظهر Lycos نتائجه ضمن أربعة أقسام تظهر جميعها على الصفحة نفسها. هذه الأقسام هي:

أ. Popular

يقوم فريق البحث في Lycos بالبحث يومياً عن المواقع الأكثر اختياراً من قبل المستخدمين بناءً على طلبات البحث المختلفة. على أساس هذه المعلومات، يقوم الفريق بربط كلمات معينة بهذه المواقع، وعندما ترد هذه الكلمات ضمن طلب بحث، تظهر هذه المواقع ضمن صفحة النتائج وضمن قسم مخصص هو Popular. يبين الشكل (6–58) هذا القسم ضمن صفحة النتائج.



الشكل (6–58): قسم **Popular** ضمن صفحة النتائج

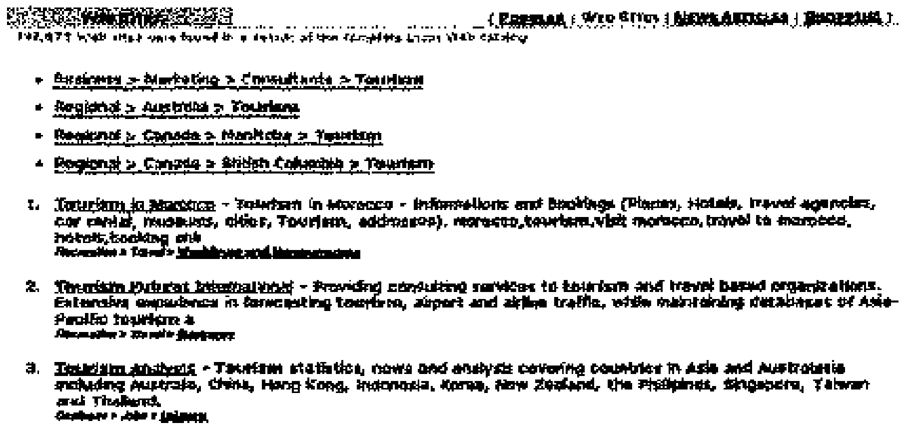
نلاحظ في أعلى هذا القسم وتحت العنوان مباشرة عدد المواقع التي وجدها النظام (وهي هنا 13). في حين تظهر كل نتيجة كارتباط تشعبي إلى ذلك الموقع، وبجانبه شرح بسيط عن الموقع، ثم URL الموقع.

ب. *مواقع الوب Web Sites*

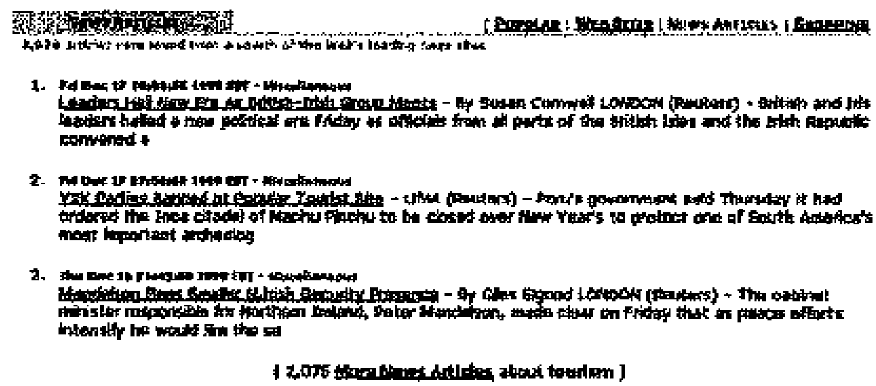
تجمع المواقع في هذا القسم من عدة مصادر، منها الصفحات التي قام المستخدمون بإضافتها إلى Lycos، والمواقع التي قام المعنيون في Lycos بإضافتها إلى مجلده يدوياً، والفئات العامة في مجلده والمطابقة لطلب المستخدم. يبين الشكل (6–59) جزءاً من هذا القسم ضمن صفحة النتائج.

ج. *مقالات الأخبار News Articles*

يحتوي هذا القسم ارتباطات تشعبية إلى مجموعة من المقالات الإخبارية الحديثة المطابقة لطلب المستخدم، كما يظهر ذلك الشكل (6–60).



الشكل (6-59): قسم Web Sites في صفحة النتائج



الشكل (6-60): قسم News Articles ضمن صفحة النتائج

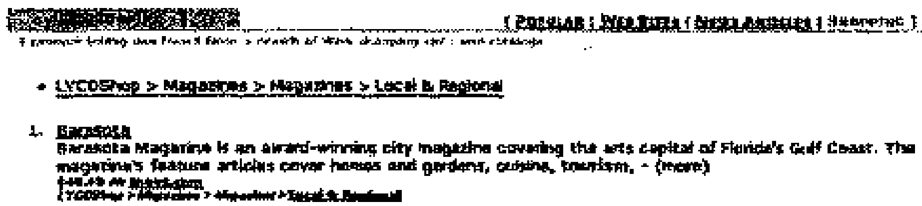
د. *التسوق Shopping*

يحتوي هذا القسم على مجموعة من الارتباطات التشعبية إلى صفحات التسوق في LYCOShop (مركز التسوق على الويب الذي يؤمنه Lycos) التي اختارها Lycos بناء على طلب المستخدم، كما يظهر ذلك الشكل (6-61).

8. لا يسمح بالتغذية الراجعة.

9. يملك Lycos قاعدة بيانات ضخمة.

10. يتمتع Lycos بسرعة جيدة، فلا تعانيْ من طول الانتظار من أجل الحصول على النتائج، رغم ضخامة قاعدة بياناته.



الشكل (6–61): قسم Shopping ضمن صفحة النتائج

ج. البحث حسب الموصفات

يدعم Lycos البحث حسب الموصفات ضمن مجموعة من أدواته مثل الصفحات الصفر والبحث عن الأشخاص والبضائع، وكنا قد تحدثنا عن هذه الأدوات في سياق حديثنا عن محركات بحث سابقة، ولا تختلف طريقة تقديم Lycos لهذه الأدوات عن محركات البحث تلك.

6–5–3 أدوات المعلومات الأخرى

1. Lycos Traffic Affiliate Program: يمكن للمستخدم الذي يملك موقعاً على الشبكة أن يشترك في هذا البرنامج ويجني من وراء اشتراكه فيه بعض المال. يكفي أن يختار واحداً أو أكثر من "صناديق المحتويات في شبكة Lycos"، ثم يضع هذه الصناديق على موقعه. تحوي هذه الصناديق على معلومات مختلفة من كافة أنحاء الشبكة. والمطلوب من المستخدم أن يختار الصناديق التي تنفع في جلب مستخدمين إلى موقعه لتفحص هذه الصناديق. ومن أجل كل مستخدم يتفحص هذه الصناديق، ـه عائد 0.02 أو 0.03 دولاراً. يوجد ارتباط تشعبي إلى هذا البرنامج على صفحة Lycos الجزء 7.

2. SearchGuard : أداة مجانية تقوم بفلترة الصفحات التي تحتوي أموراً يجب على الأطفال عدم الاطلاع عليها. تُشبه هذه الأداة من ناحية وظيفتها الأدوات التي كنا قد رأيناها سابقاً في AltaVista وInfoSeek (Family Filter, GOguardian). يمكن الوصول إليها من الارتباط التشعبي Parantal Control أسفل حقل الحوار المخصص للبحث.

6-5-4 إضافة موقع إلى محرك البحث Lycos

يمكن للمستخدم أن يضيف موقعه إلى Lycos باتباع الارتباط التشعبي على صفحة Lycos الرئيسية ضمن الجزء (Add your site to Lycos). كل ما على المستخدم عمله هو إعطاء عنوان موقعه، وعنوان E-mail الخاص به. يقوم عنكبوت Lycos بالدخول إلى الموقع والتصفح فيه من أجل فهرسة صفحاته، وقد يحتاج الأمر إلى أسبوعين أو ثلاثة أسابيع لإضافة الموقع إلى قاعدة بيانات Lycos.

يُلزم Lycos المستخدم عند إضافة موقعه بأن لا يحتوي URL هذا الموقع أحد الرموز التالية: (&, %, ?, $, =)، وإلا لن يتمكن عنكبوته من التصفح والتعرف عليه.

6-6 محرك البحث Excite

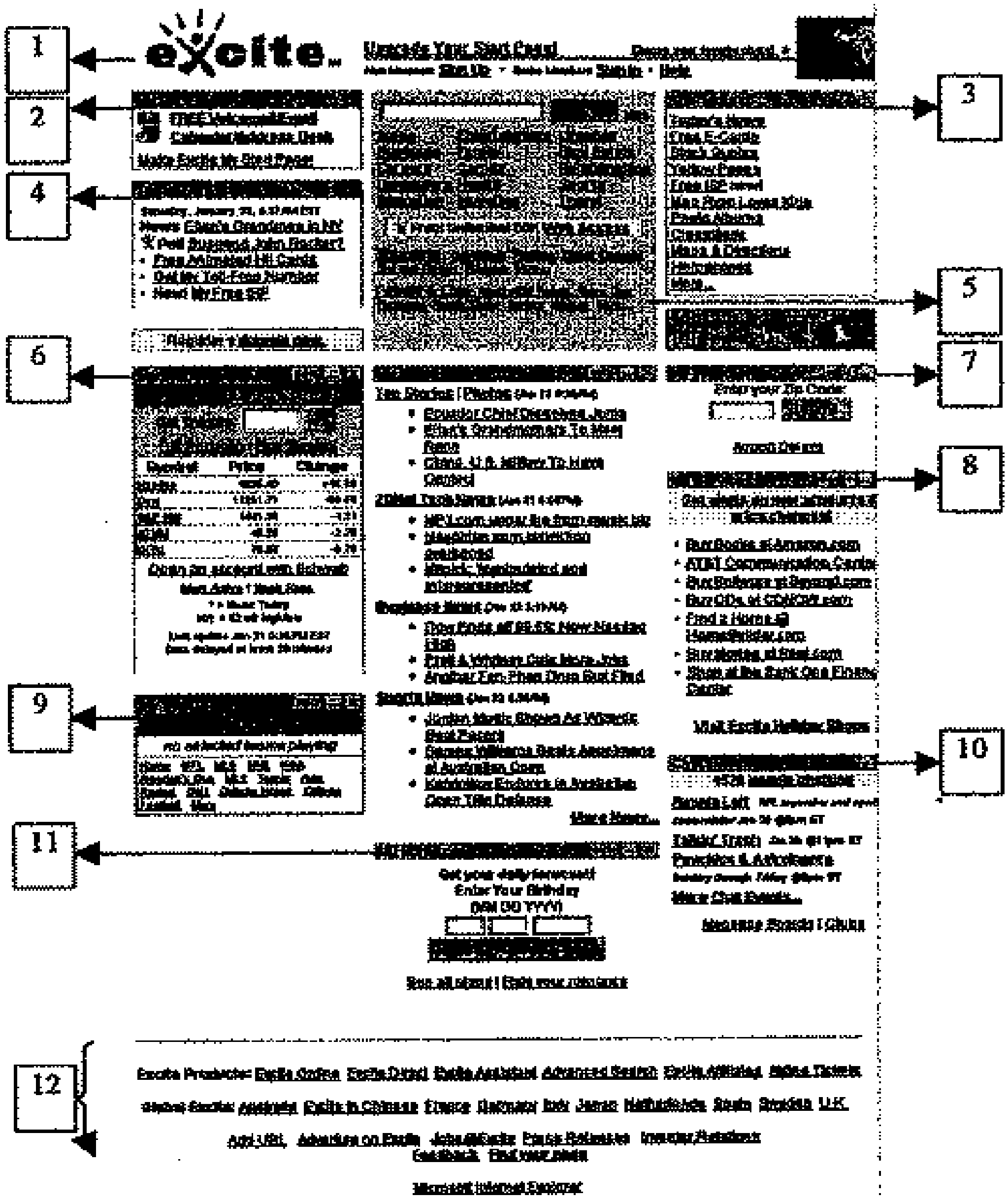
WWW.Excite.com

يُعتبر Excite من محركات البحث الشعبية والجيدة للبحث في الويب، فلديه فهرسٌ بـ 50 مليون صفحة، كما أنه يحتل مرتبة هامة بين محركات البحث من حيث عدد الزوار. وتتضمن إجاباته غالباً معلومات عن الشركات وبعض النتائج الرياضية في الحالات التي يجد فيها هذه المعلومات مناسبة لطلب المستخدم.

أُطلق Excite في أواخر عام 1995 وتطور بسرعة كبيرة، بحيث سبق منافسين رئيسيَّين هما: Magellan في تموز 1996 وwebCrawler في تشرين ثاني 1996

6-6-1 واجهة Excite

يُظهر الشكل (6-62) صفحة Excite على الوب، والتي يمكن الوصول إليها باستخدام العنوان: http:\\www.excite.com، نجد فيما يلي الإمكانات المختلفة لـ Excite:



الشكل (6-62): صفحة Excite

- الجزء 1: رمز Excite.
- الجزء 2: ارتباط تشعبي إلى My Excite ، وهو صفحة Excite التي يمكن تشكيلها حسب رغبة المستخدم.
- الجزء 3: مجموعة من أدوات Excite.
- الجزء 4: مجموعة من الاخبار.
- الجزء 5: مكان صياغة طلبات البحث، ومجلد Excite.
- الجزء 6: أمور مالية.
- الجزء 7: مكان الحصول على النشرة الجوية.
- الجزء 8: مجموعة من الخدمات.
- الجزء 9: قسم مخصص للرياضة.
- الجزء 10: مجموعة من جلسات الحوار على الإنترنت (Chat).
- الجزء 11: قسم مخصص للأبراج.
- الجزء 12: ارتباطات تشعبية إلى مجموعة من منتجات Excite ومواقع Excite في العالم.

6-6-2 البحث عن المعلومات في Excite

يتيح Excite الأنواع الثلاثة من طرق البحث عن المعلومات: التصفح، البحث عن المحتوى، البحث عن الموصفات.

آ. التصفح

يشابه البحث بالتصفح في Excite التصفح في باقي محركات البحث، فهو يمتلك مجلداً خاصاً تُصنَّف المواقع ضمنه حسب مواضيعها. ويمكن التصفح فيه باستخدام الارتباطات التشعبية ضمن الجزء [5].

ب. البحث عن المحتوى

يتيح Excite إمكانية صياغة طلبات البحث من خلال الجزء [5] من صفحته الرئيسية. كما يتيح واجهات بحث أخرى من خلال الارتباط التشعبي More الموجود إلى يمين الزر Search في الجزء [5]. عند اتباع هذا الارتباط التشعبي نحصل على الواجهة الموضحة في الشكل (6–63) التي تحتوي الإمكانات التالية:

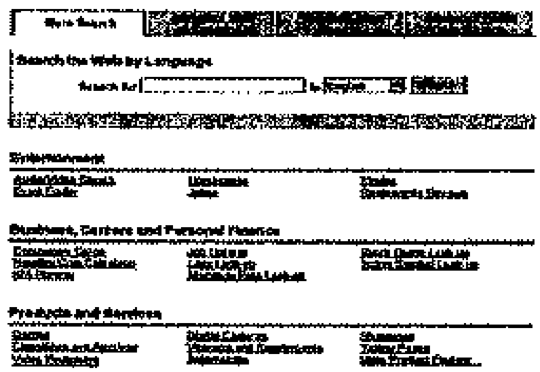


الشكل (6–63): واجهات البحث المختلفة التي تظهر باتباع الارتباط التشعبي More

أ. More Search: يسمح بالبحث باستخدام الفلترة حسب اللغة، كما يصنف أدوات Excite المختلفة ضمن فئات ويظهرها للمستخدم الذي يستطيع استخدام الأداة التي يريدها من الفئة التي يندرج طلبه في إطارها. يُظهر الشكل (6–64) جزءاً من هذه الواجهة.

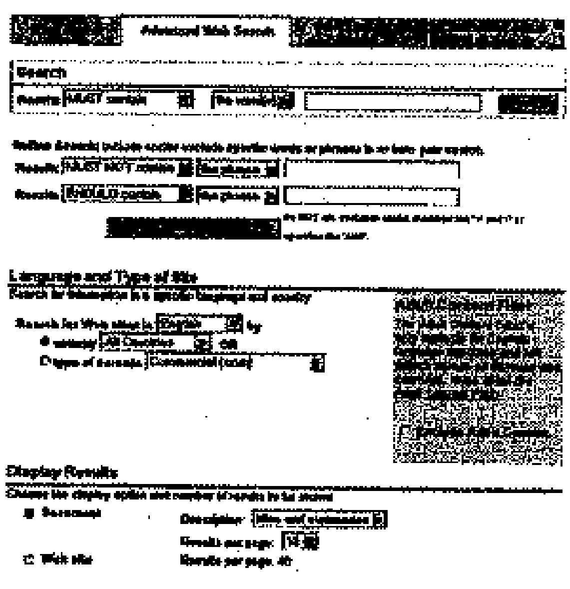
ب. Advanced Web Search: يسمح بصياغة طلبات بحث متقدم دون الحاجة إلى معرفة قواعد كتابة الطلبات المتقدمة، وذلك من خلال المنزلقات المختلفة التي تسمح بتحديد ضرورة ورود الكلمة (الجملة) أو عدم ورودها ضمن الصفحات المطابقة، ونوعية الطلب: أهو جملة أم كلمة، كما يُظهر ذلك الشكل (6–65).

في حين يسمح القسم Language and Type of Site بتحديد فلترة على لغة الصفحات والبلد الذي توجد فيه الصفحات وDomain Name للمواقع التي سيبحث ضمنها. أما القسم Display Results، فيحدد خياراتٍ لإظهار النتائج وعددها.

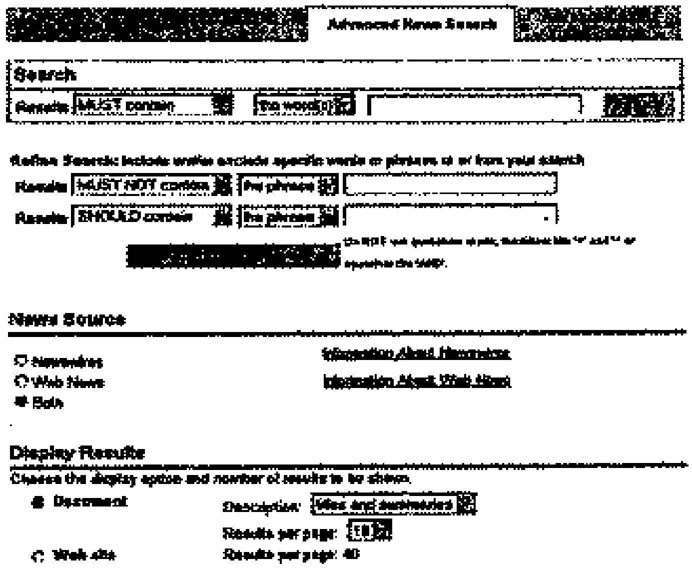


الشكل (6-64): واجهة More Search في Excite

ج. Advanced News Search: يسمح بالبحث عن مقالات الأخبار المطابقة لطلب المستخدم كما يبين ذلك الشكل (6-66). إن طريقة صياغة الطلب هنا هي نفسها المتبعة مع Advanced Web Search، ونلاحظ وجود القسمين التاليين: القسم News Source الذي يسمح بتحديد مصدر المقالات الذي نريد البحث ضمنه، والقسم Display Results الذي سبق وتحدثنا عنه من أجل Advanced Web Search.

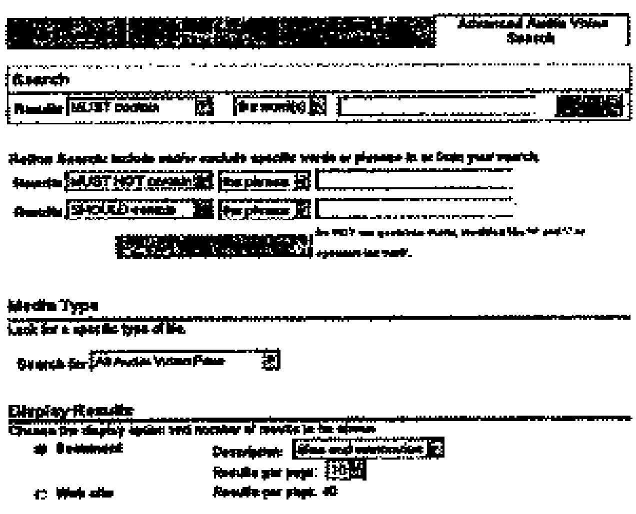


الشكل (6-65): واجهة Advanced Web Search في Excite



الشكل (6–66): واجهة Advanced News Search في Excite

د. Advanced Audio Video Search: يسمح بالبحث عن مختلف أنواع ملفات الصوت والفيديو كما يبين ذلك الشكل (6–67).



الشكل (6–67): واجهة Advanced News Search في Excite

تشابه طريقةُ صياغة طلبات البحث هنا أيضاً طريقةَ صياغتها في Advanced Web Search. ونلاحظ وجود قسمين ضمن الصفحة هما: Media Type الذي يسمح بتحديد فلترة على الملفات المعادة كي تكون من نمط معين، و Display Results التي ذكرناها آنفاً.

تقييم البحث بالمحتوى

1. البحث البسيط

يسمح Excite بعمليات البحث البسيط، ومن أجل فحص إمكانات هذا النوع من البحث، قمنا بالتجارب التالية:

- التحسس للأحرف الصغيرة والكبيرة: بالبحث عن tourism و Tourism، نجد اختلافاً في عدد صفحات المعادة كلَّ مرة. مما يعني أنه يتحسس للأحرف الكبيرة والصغيرة. وهو يتحسس للأحرف على النحو الذي يتحسس فيه Lycos لها.
- البحث عن جمل: بالبحث عن "tourism in Syria"، نجد أن الصفحات التي وجدها النظام تتحدث عن السياحة في سورية، وتحوي هذه الجملة. فهو إذن يتيح إمكانية تجميع كلمات البحث في جمل، إذ يكفي وضع الجملة ضمن "" كي يقوم بعملية بحث عن الصفحات التي تحوي الجملة بكاملها.
- الكلمات المفتاحية اللازمة: يسمح Excite للمستخدم بأن يحدد كلمة يريد لها أن تكون محتواة ضمن جميع الصفحات المطابقة. فمثلاً يؤدي الطلب: tourism +Syria إلى إيجاد الصفحات التي تتحدث عن tourism وتحتوي ضمنها بالضرورة كلمة Syria.
- الكلمات المفتاحية الممنوعة: يتيح Excite إمكانية منع كلمة من أن تكون موجودة ضمن صفحة ما عن طريق استخدام "-". فمثلاً: بالبحث عن "tourism" والبحث عن "tourism -Syria"، نجد بفحص النتائج أن الصفحات التي تتحدث عن "Syria" قد اختفت من نتائج الطلب الثاني.

- لا يسمح باستخدام الرموز العامة (Wild Card).
- يرتب النتائج حسب درجة المطابقة المتناقصة التي يحسبها باستخدام تابع مطابقة خاص، ويمكن ترتيبها حسب الموقع الذي توجد فيه.

2. البحث المتقدم

يسمح Excite باستخدام البحث المتقدم من خلال السماح باستخدام العوامل البوليانية: AND, OR, AND NOT.

3. الفلترة

يسمح Excite عند استخدام Advanced Web Search بالفلترة حسب اللغة والمنطقة الجغرافية ومجالات الأسماء على الإنترنت (Domain Names)، في حين يتيح الفلترة حسب اللغة فقط عند استخدام More Search، أما في حالة Advanced Audio Video Search، فيمكن الفلترة حسب نمط الوسائط (AVI, MPEG, ...). إن الفلترة، كما نلاحظ، بسيطة في Excite وتتعلق بواجهة البحث المستخدمة.

5. يؤمن Excite تصنيفاً جيداً للمواضيع ضمن فئات بفضل مجلده المبني يدوياً، ونجد الفئات الرئيسية لمجلده ضمن الجزء [5] من الصفحة الرئيسية.

6. إظهار النتائج

يُظهر Excite نتائجه ضمن أربعة أقسام تظهر جميعها على الصفحة نفسها كما يبين ذلك الشكل (6–68). هذه الأقسام هي:

أ. طلبات مقترحة من Excite (Did You Mean?)

يقترح Excite كلمات بحث لإضافتها إلى طلب البحث الأصلي، وهكذا يستطيع المستخدم أن يختار من هذه الكلمات المجموعة التي توافقه، ويكرر طلب البحث. وقد رأينا سابقاً بعض محركات البحث تقوم بعمل مشابه.

ب. نتائج البحث في مجلد Excite (Directory)

وهي الفئات التي وجدها Excite مطابقةً لطلب المستخدم ضمن مجلده الخاص. يعرض Excite بجانب هذه الفئات مجموعة من الكتب التي تطابق طلب المستخدم أيضاً.

ج. نتائج البحث في صفحات الوب (Web search results)

وهي الصفحات المطابقة لطلب المستخدم والتي وجدها Excite على الوب. نلاحظ في هذه النتائج ما يلي:

- يعرض Excite عدد الصفحات المطابقة التي وجدها، وهو هنا 323720 صفحة.
- يعرض Excite النتائج، مع الشرح أو بدونه (الارتباط التشعبي Titles only)، مرتبة حسب درجة المطابقة أو حسب الموقع الذي يحتويها (الارتباط التشعبي Web Site).
- يعرض Excite كلّ نتيجة كما يلي:
  - ارتباط تشعبي إلى الصفحة النتيجة، تسبقه درجة المطابقة التي حسبها Excite وفق خوارزمياته الخاصة، ونجد أن درجة مطابقة أول نتيجة معروضة تساوي 64٪.
  - تحت العنوان، يوجد شرح بسيط عن هذه الصفحة، وفي نهايته URL هذه الصفحة.
  - ارتباط تشعبي Search for more documents like this، وهو يؤدي إلى تنفيذ بحث جديد عن صفحات مشابهة للصفحة النتيجة الحالية.

د. *نتائج البحث في الأخبار*

ويعرض Excite النتائج فيها كما يلي:

- عدد النتائج التي وجدها، وهي هنا 367.
- إمكانية إظهار النتائج مع الشرح أو بدونه (الارتباط التشعبي Summaries)، وترتيب النتائج حسب النشر (Publication) أو التاريخ (Date).
- عرض كل نتيجة كارتباط تشعبي إلى صفحة الأخبار المعنية، وبجانبها مصدر الخبر، ثم ارتباط تشعبي إلى مجموعة من المقالات المتمثلة به.

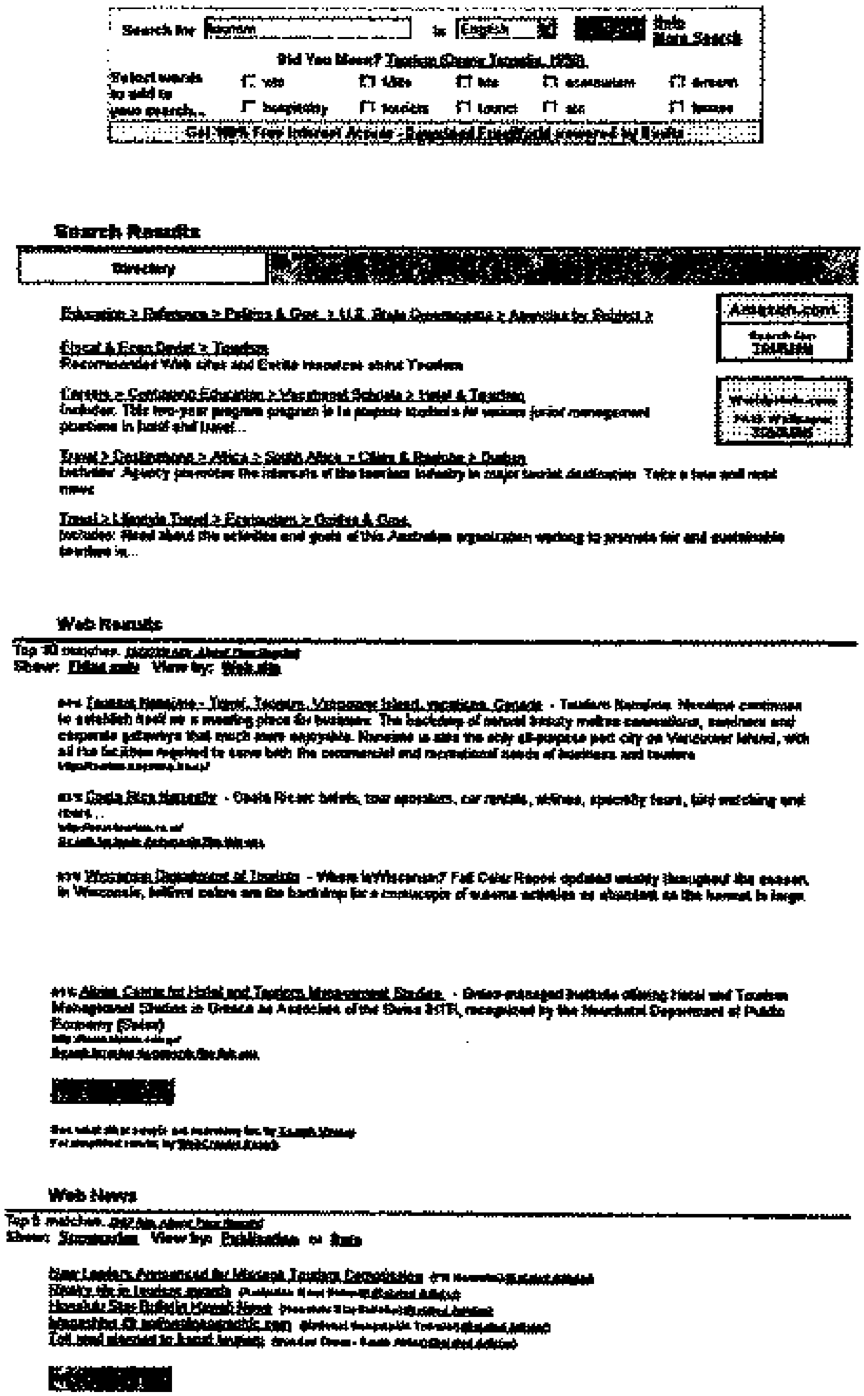
7. لا يسمح بالتغذية الراجعة.

8. يملك Excite قاعدة بيانات ضخمة.

9. يتمتع Excite بسرعة جيدة، فلا نعاني من طول الانتظار من أجل الحصول على النتائج، رغم ضخامة قاعدة بياناته.

ج. البحث حسب الموصفات

يدعم Excite البحث حسب الموصفات ضمن مجموعة من أدواته مثل الصفحات البيض، والصفحات الصفر، والأبراج، وغيرها من الأدوات التي نجدها على الصفحة الرئيسية.



الشكل (6-68): صفحة النتائج في Excite

6-6-3 إضافة موقع إلى محرك البحث Excite

يمكن إضافة موقع إلى Excite من خلال الارتباط التشعبي Add a URL الموجود في الجزء 6 من الصفحة الرئيسية. يطلب Excite من المستخدم أن يعطيه URL موقعه، كما يطلب معلومات إضافية مثل E-mail واللغة الأساسية المستخدمة، والموقع الجغرافي، إضافة إلى فئة الموقع المقترحة. يقوم عنكبوت Excite بعدئذ بزيارة الموقع، وفهرسة الصفحات التي ضمنه.

يحتاج Excite إلى حوالي أسبوعين حتى يضيف الموقع إلى فهرسه، ويمكن منع Excite من فهرسة صفحات معينة باستخدام الملف Robots.txt (كما في محركات البحث السابقة).

ننوه أخيراً إلى عدم إمكانية استخدام مؤشرات HTML الخاصة (كما هو الحال مع محركات بحث أخرى) في عملية توجيه عنكبوت Excite، أو من أجل إضافة كلمات مفتاحية. عندما يجد Excite مؤشرات HTML خاصة، فإنه يتجاهلها كما لو كانت غير موجودة. الاستثناء الوحيد هو في إمكانية إضافة شرح للصفحة، فهو المؤشر الوحيد الذي يأخذه Excite بعين الاعتبار.

6-7 محرك البحث WebCrawler

"It's that Simple."

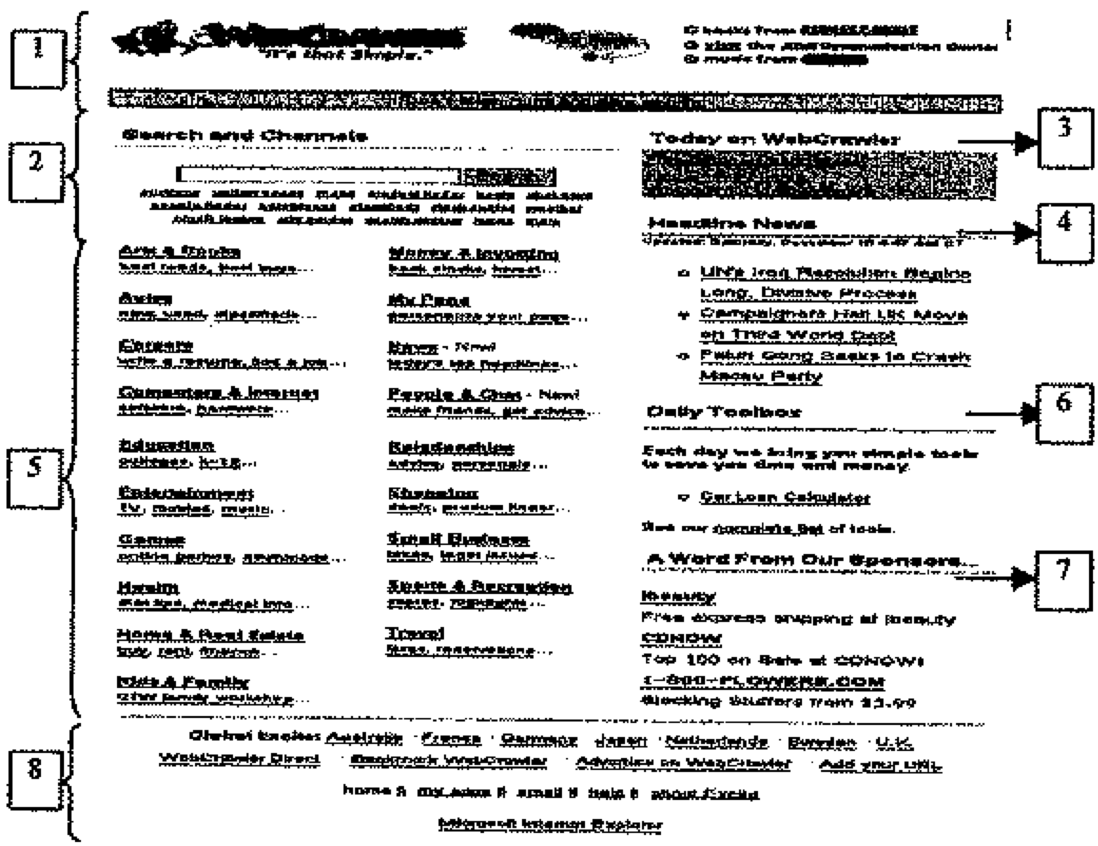
WWW.WebCrawler.com

انطلق محرك البحث WebCrawler كمشروع طلابي في جامعة واشنطن في عام 1994، أي بعد ظهور الوب بقليل، واستثمر من قبل الجامعة لفترة من الزمن قبل أن تشتريه America Online ليصبح مشروعاً تجارياً. بعد ذلك، ضمته شبكة Excite ليصبح إحدى دعاماتها الأساسية.

يُعتبر WebCrawler من محركات البحث الضخمة، فهو يمتلك قاعدة بيانات تحتوي 50 مليون صفحة، ويعطي أفضل النتائج عند الإجابة على طلبات البحث عن المعلومات المكتوبة بلغة طبيعية أو باستخدام التعابير البوليانية. وباشتراكه ضمن شبكة Excite، ومع تزايد الشركاء، فإنه يؤمن المزيد من الأدوات المساعدة والخدمات على الوب التي تُغنيه وتنميه.

6-7-1 واجهة WebCrawler

يُظهر الشكل (6-69) صفحة WebCrawler على الوب، والتي يمكن الوصول إليها باستخدام العنوان: http:\\www. WebCrawler.com. نجد فيما يلي الإمكانات المختلفة لـ WebCrawler :



الشكل (6-69): صفحة WebCrawler

- الجزء 1: شعار WebCrawler.
- الجزء 2: يسمح للمستخدم بالقيام بعمليات بحث.
- الجزء 3: آخر أخبار WebCrawler.
- الجزء 4: عناوين الأخبار.
- الجزء 5: مجلد WebCrawler.
- الجزء 6: الأداة اليومية من أدوات WebCrawler.
- الجزء 7: مجموعة من الإعلانات لأدوات على الإنترنت.
- الجزء 8: ارتباطات تشعبية لزيارة WebCrawler في دول مختلفة، إضافة إلى مجموعة أخرى من أدوات WebCrawler.

6-7-2 البحث عن المعلومات في WebCrawler

يتيح WebCrawler الأنواع الثلاثة من طرق البحث عن المعلومات: التصفح، البحث عن المحتوى، البحث عن الموصفات.

آ. التصفح

يشابه البحث بالتصفح في WebCrawler التصفح في باقي محركات البحث، فهو يمتلك مجلداً خاصاً تُصنَّف المواقع ضمنه حسب مواضيعها. ويمكن التصفح فيه باستخدام الارتباطات التشعبية ضمن الجزء 5.

ب. البحث عن المحتوى

يتيح WebCrawler إمكانية صياغة طلبات البحث من خلال الجزء [5] من صفحته الرئيسية.

تقييم البحث بالمحتوى

1. البحث البسيط

يسمح WebCrawler بعمليات البحث البسيط، ومن أجل فحص إمكانات هذا النوع من البحث، قمنا بالتجارب التالية:

- التحسس للأحرف الصغيرة والكبيرة: بالبحث عن tourism و Tourism، لا نجد اختلافاً في عدد صفحات المعادة كلّ مرة. مما يعني أنه لا يتحسس للأحرف الكبيرة والصغيرة.
- البحث عن جمل: بالبحث عن "tourism in Syria"، نجد أن الصفحات التي وجدها النظام تتحدث عن السياحة في سورية، وتحوي هذه الجملة. فهو إذن يتيح إمكانية تجميع كلمات البحث في جمل، إذ يكفي وضع الجملة ضمن "" كي يقوم بعملية بحث عن الصفحات التي تحوي الجملة بكاملها.
- الكلمات المفتاحية اللازمة: يسمح WebCrawler للمستخدم بأن يحدد كلمة يريد لها أن تكون محتواة ضمن جميع الصفحات المطابقة. فمثلاً يؤدي الطلب: tourism +Syria إلى إيجاد الصفحات التي تتحدث عن tourism وتحتوي ضمنها بالضرورة كلمة Syria.
- الكلمات المفتاحية الممنوعة: يتيح WebCrawler إمكانية منع كلمة من أن تكون موجودة ضمن صفحة ما عن طريق استخدام "-". فمثلاً: بالبحث عن "tourism" والبحث عن "- tourism Syria"، نجد بفحص النتائج أن الصفحات التي تتحدث عن "Syria" قد اختفت من نتائج الطلب الثاني.
- لا يسمح WebCrawler باستخدام الرموز العامة (Wild Card).

- لا يسمح WebCrawler بالبحث المتقدم.

2. البحث البولياني

يُتيح WebCrawler القيام بالبحث البولياني باستخدام العوامل البوليانية : AND, OR, NOT. يسمح WebCrawler بصياغة طلبات مثل:

(Tourism AND Syria) AND (NOT Aleppo)

من أجل البحث عن الصفحات التي تتحدث عن السياحة في سورية عدا تلك التي تتحدث عن حلب.

3. لا يسمح WebCrawler بالفلترة.

4. يؤمن WebCrawler تصنيفاً جيداً للمواضيع ضمن فئات بفضل مجلده المبني يدوياً، وتجد الفئات الرئيسية لمجلده ضمن الجزء 5 من الصفحة الرئيسية.

5. إظهار النتائج

يُظهر WebCrawler نتائجه ضمن خمسة أقسام تظهر جميعها على الصفحة نفسها كما يبين ذلك الشكل (6-70). هذه الأقسام هي:

- في أعلى الصفحة، نجد حقل حوار لصياغة طلبات بحث جديدة، أو تعديل الطلب السابق.
- نتائج البحث في الوب (Web Results): يُعطي WebCrawler عدد الصفحات المطابقة لطلب المستخدم، والتي وجدها على الوب (في مثالنا عددها 23643). وهو يعرض عنوان الصفحة النتيجة فقط، ويمكن أن يطلب المستخدم عرض شرح عن الصفحات النتائج باستخدام الارتباط التشعبي summaries. في الحالات التي لا يكون فيها للصفحة عنوان، يكتفي WebCrawler بكتابة URL الصفحة فقط.

- الاختصارات (Shortcuts): وهي نتائجٌ مطابقةٌ لطلب المستخدم من مصادر أخرى. وعلى الشكل، نجد اختصارات إلى مجموعة المجموعات الأخبارية Usenet، وإلى مجموعات تتعلق بالسياحة (tourism)، كما نجد اختصاراً إلى فئة من فئات WebCrawler وهي Travel.
- أخبار مطابقة لطلب المستخدم News Stories about: وهي نتائج البحث ضمن الأخبار News.
- أخيراً وفي أسفل الصفحة، نجد ارتباطاً تشعبياً إلى Excite من أجل القيام بنفس البحث عليه.

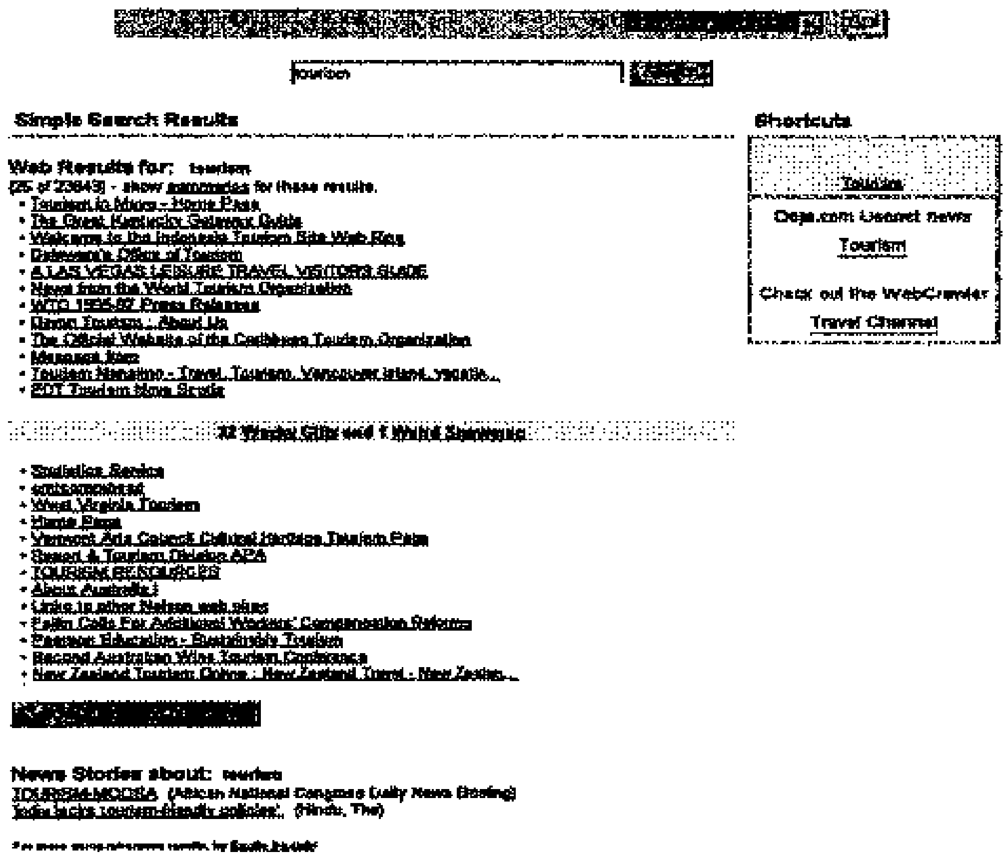
6. لا يسمح بالتغذية الراجعة.

7. يملك WebCrawler قاعدة بيانات ضخمة.

8. يتمتع WebCrawler بسرعة جيدة، فلا نعاني من طول الانتظار من أجل الحصول على النتائج، رغم ضخامة قاعدة بياناته.

ج. البحث حسب الموصفات

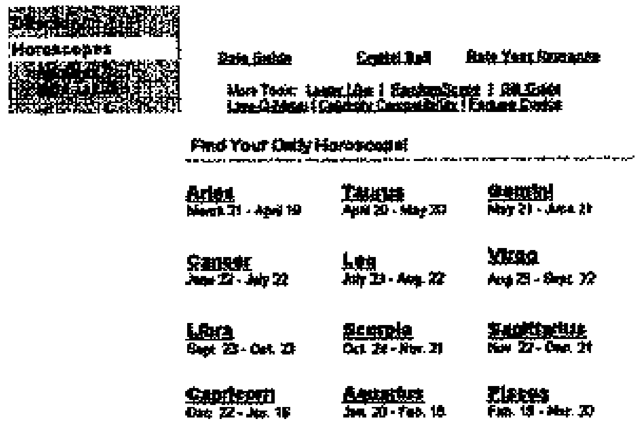
يدعم WebCrawler البحث حسب الموصفات ضمن مجموعة من أدواته مثل الصفحات الصفر وأداة البحث عن الأشخاص People finder، وغيرها... وقد تحدثنا سابقاً عن هذه الأدوات.



الشكل (6–70): صفحة النتائج في WebCrawler

6–7–3 أدوات المعلومات الأخرى

تسمح أداة الأبراج (Horoscopes) لمستخدمي WebCrawler بأن يلقوا نظرة علـى مـا تقولـه الأبـراج لهم كل يوم، ونجد في الشكل (6–71) كيف يعرض الأبراج كارتباطات تشـعبية إلى الصفحـات الـتي تحتوي التوقعات.



الشكل (6–71): أداة الأبراج في WebCrawler

6–7–4 إضافة موقع إلى محرك البحث WebCrawler

لا يحتاج WebCrawler إلى أكثر من URL الموقع، وURL الصفحات الجانبية المساعدة في الموقع كما يُظهر ذلك الشكل (6–72). وهو، بعكس بعض محركات البحث، لا يفهرس كامل صفحات الموقع، لكنه يفهرس الصفحات التي يختارها وفق طريقة خاصة بحيث يؤمن للمستخدم الصفحات التي يمكن أن تحتوي معلومات مفيدة. إلا أن WebCrawler يزور جميع الصفحات ويحفظ الـ URL الخاصة بها من أجل زياراته اللاحقة والتحقق من التعديلات عليها.

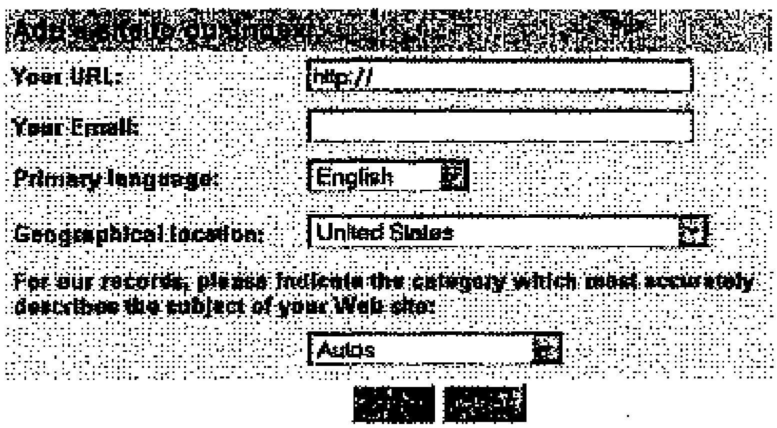
يضع WebCrawler شرطاً على عدد الصفحات التي يمكن للمستخدم أن يسجلها ضمن الموقع وهو 25 صفحة، وعند تجاوز هذا الرقم، يقوم أخصائيو WebCrawler بزيارة ذلك الموقع، ولن يُضاف الموقع إلى فهرس WebCrawler إلاّ إذا وجد أولئك الأخصائيون بأن الموقع جدير بأن يُضاف.

لا يطلب WebCrawler من المستخدم أن يُرفق مع URL موقعه أي كلمات مفتاحية أو فئات مقترحة، فهو يقوم بفهرسة نص الصفحة كاملاً شرطَ أن تكون الصفحة ذات حجم أصغر من واحد ميغا بايت.

يجب على المستخدم أن يحذر من محاولات غش WebCrawler، لأن WebCrawler سيلغي كامل الموقع من فهرسه حالَ اكتشاف هذه المحاولات. أما الطريقة التي ينصح WebCrawler المستخدمين

باتباعها من أجل تحسين ترتيب مواقعهم، فهي بإعطاء الصفحات عناوين دقيقة تحتوي كلمات محددة، كما ينصحهم بأن تحتوي صفحة الموقع الأساسية مجموعة الكلمات والمواضيع الموجودة في الموقع، لأن WebCrawler يستفيد من صفحة الموقع ومن العناوين في عملية الفهرسة.

أخيراً، نذكر أن WebCrawler يحتاج إلى أسبوعين لإضافة موقع.



الشكل (6-72): إضافة موقع إلى WebCrawler

الفصل السابع

محركات البحث البينية والعربية

7-1 محركات البحث البينية

إن السمة الأساسية لمحركات البحث البينية هي أنها لا تملك قاعدة بيانات خاصة بها، بل تستعين بقواعد بيانات محركات البحث الأخرى من أجل الإجابة على طلبات المستخدمين. تكمن الفكرة وراء هذا العمل في قصور أي محرك بحث لوحده عن ملاحقة وتتبع جميع الصفحات الجديدة والتغييرات التي تطرأ على الوب، كما أن كل محرك يختلف عن غيره في الصفحات التي يفهرسها والتغييرات التي يرصدها. إذن يكمن الحل في استخدام قواعد بيانات أفضل محركات البحث، كي تزداد إمكانية ملاحقة التغييرات التي رصدتها محركات البحث هذه على الوب.

يعتبر القائمون على محركات البحث البينية أن استخدام محركات البحث من هذا النوع له نواحٍ إيجابية متعددة، منها:

- إمكانية الحصول على معلومات حديثة أكثر مما يوفّره غيرها من محركات البحث كلٌ على حدة.

- لن يحتاج المستخدم لتعلم طرق صياغة الطلبات إلا من أجل محرك البحث البيني الذي يستخدمه، ورغم ذلك، يستطيع توجيه الطلب إلى أي محرك بحث من المحركات المتوفرة (أو إلى مجموعة منها).

اخترنا من بين العديد من محركات البحث البينية المعروفة دراسة أشهر محركي بحث بيني: MetaCrawler, theBigHub.

7-2 محرك البحث البيني MetaCrawler

WWW.MetaCrawler.com

نشأ MetaCrawler في جامعة واشنطن نتيجة مشروع طلابي، ووُضع على الوب عام 1995. بعد ذلك أصبح تابعاً لشركة Go2Net التي تحرص على أن يبقى أسرع المحركات البينية الموجودة على الوب.

يستقبل MetaCrawler طلبات البحث كطلبات HTTP، ثم يحول هذه الطلبات لتلائم صيغة طلبات محركات البحث التي سيرسلها إليها. ينتظر MetaCrawler حتى تأتيه الإجابات من جميع محركات البحث، أو حتى ينتهي الزمن المحدد، ثم يعرض هذه الإجابات للمستخدم بعد حذف الإجابات المكررة وترتيبها حسب مطابقتها للطلب.

7-2-1 واجهة MetaCrawler

يُظهر الشكل (7-1) صفحة MetaCrawler على الوب، والتي يمكن الوصول إليها باستخدام العنوان: http:\\www. MetaCrawler.com. نجد فيما يلي الإمكانات المختلفة لـ MetaCrawler:

الجزء 1: شعار MetaCrawler.

الجزء 2: واجهة البحث، مع جميع الارتباطات إلى واجهات البحث الأخرى.

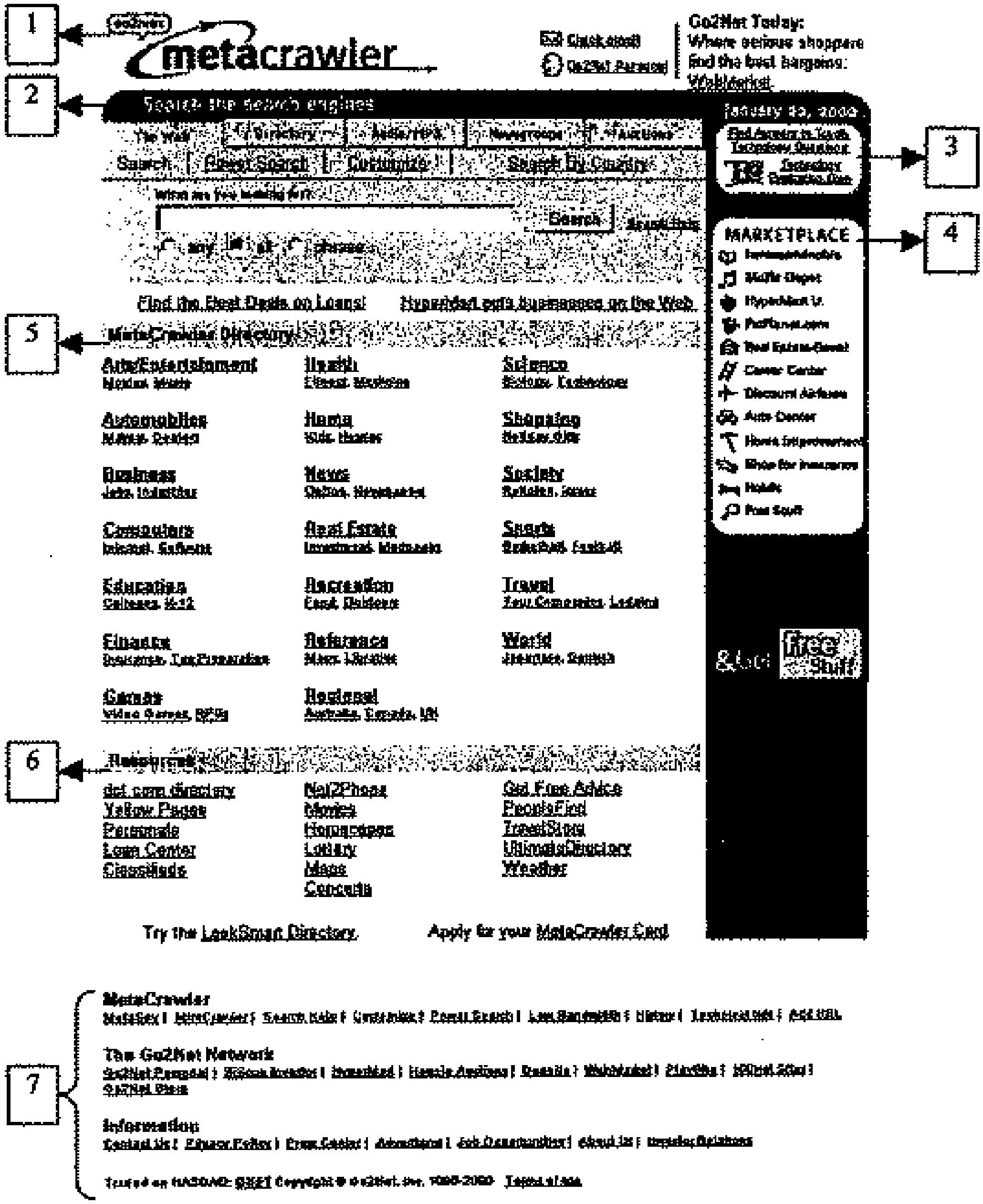
الجزء 3: خدمة للإجابة عن الأسئلة والاستفسارات التكنولوجية.

الجزء 4: التسوق في MetaCrawler.

الجزء 5: مجلد MetaCrawler.

الجزء 6: مجموعة أدوات MetaCrawler.

الجزء 7: ارتباطات إلى Go2Net وأدوات MetaCrawler.



الشكل (7-1): صفحة MetaCrawler

7-2-2 البحث عن المعلومات في MetaCrawler

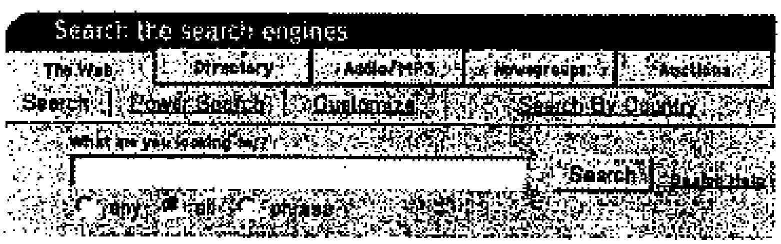
يتيح MetaCrawler الأنواع الثلاثة من طرق البحث عن المعلومات: التصفح، البحث عن المحتوى، البحث عن الموصفات.

آ. التصفح

يشابه البحث بالتصفح في MetaCrawler التصفح في باقي محركات البحث، وباعتبار أنه لا يوجد لـ MetaCrawler قاعدة بيانات خاصة، فهو يستخدم المجلد المفتوح Open Directory. يظهر المجلد في الشكل (7-1) ضمن الجزء 5.

ب. البحث عن المحتوى

يتيح MetaCrawler إمكانية صياغة طلبات البحث من خلال الجزء 2 من صفحته الرئيسية. يظهر الشكل (7-2) واجهة البحث هذه، حيث يقسم البحث إلى خمسة أنواع هي:



الشكل (7-2): واجهة البحث في الصفحة الرئيسية لـ MetaCrawler

- The Web: للبحث عن صفحات الوب.
- Directory: للبحث ضمن المجلد المفتوح.
- Audio/MP3: للبحث عن ملفات الصوت.
- Newsgroups: للبحث في مجموعات الأخبار.

- Auctions: البحث في أكثر من 30 موقعا للمزادات عن بضائع معينة، منها تلك الموجودة على Yahoo! وeBay وAmazone.com وMSN وغيرها....

تشترك هذه الأنواع الخمسة بشكل واجهة البحث، مع بعض الخصوصية لـ The Web، أما عناصر واجهة البحث فهي:

- الحقل الحواري وهو مكان صياغة طلبات المستخدم.
- مجموعة الخيارات في أسفل مربع الحوار، وهي تحدد طريقة معاملة الطلب:
  - any: استحضار الصفحات[1] التي تحتوي أيا من كلمات البحث.
  - all: استحضار الصفحات التي تحتوي جميع كلمات البحث. وإذا لم يجدها جميعها، فإنه يعيد الصفحات التي تحتوي العدد الأكبر من كلمات البحث.
  - phrase: إيجاد الصفحات التي تحتوي هذه الكلمات معا. في معظم الأحيان يعيد MetaCrawler صفحات تحوي جملة تقريبية، ولا يعيد دوما صفحات تحتوي الجملة تماما.

وفيما يلي سنشرح النوع The Web لأن فيه بعض الخصوصية. تتضمن واجهة البحث ضمن الوب مجموعة من الارتباطات التشعبية هي:

- Power Search: يسمح لنا بتحديد مجموعة المحركات التي نريدها أن تشترك في البحث، كما يستطيع فلترة النتائج حسب المكان الذي تنتمي إليه، وزمن انتظار كل محرك بحث، كما يحدد عدد النتائج التي تظهر في كل صفحة، وعدد النتائج المعادة من كل محرك بحث، وأخيرا طريقة ترتيب النتائج (حسب المطابقة أو المكان الذي تنتمي إليه أو محرك البحث).
- Customize: يسمح بتحديد الخيارات التي يريد المستخدم استعمالها في عمليات البحث.

---

[1] كلمة صفحات تتضمن ملفات الصوت عند البحث باستخدام *Audio/MP3*.

- Search by Country : يسمح هذا الارتباط بتحديد مجموعة من دول العالم نريد للصفحات الأجوبة أن تعاد منها. فمثلا تحديد France وGermany سيختصر نتائج البحث على الصفحات الموجودة على مخدمات ضمن هذين البلدين.

تقييم البحث بالمحتوى

1. البحث البسيط

يسمح MetaCrawler بعمليات البحث البسيط، ومن أجل فحص إمكانات هذا النوع من البحث، قمنا بالتجارب التالية:

* التحسس للأحرف الصغيرة والكبيرة: بالبحث عن tourism و Tourism، لا نجد اختلافا في عدد صفحات المعادة كل مرة. مما يعني أنه لا يتحسس للأحرف الكبيرة والصغيرة.
* البحث عن جمل: بالبحث عن "tourism in Syria"، نجد أن الصفحات التي وجدها النظام تتحدث عن السياحة في سورية، وتحوي هذه الجملة. فهو إذن يتيح إمكانية تجميع كلمات البحث في جمل، إذ يكفي وضع الجملة ضمن "" كي يقوم بعملية بحث عن الصفحات التي تحوي الجملة بكاملها.
* الكلمات المفتاحية اللازمة: يسمح MetaCrawler للمستخدم بأن يحدد كلمة يريد لها أن تكون محتواة ضمن جميع الصفحات المطابقة. فمثلا يؤدي الطلب: tourism +Syria إلى إيجاد الصفحات التي تتحدث عن tourism وتحتوي ضمنها بالضرورة كلمة Syria.
* الكلمات المفتاحية الممنوعة: يتيح MetaCrawler إمكانية منع كلمة من أن تكون موجودة ضمن صفحة ما عن طريق استخدام "-". فمثلا: بالبحث عن "tourism" والبحث عن "tourism -Syria"، نجد بفحص النتائج أن الصفحات التي تتحدث عن "Syria" قد اختفت من إجابات الطلب الثاني.

* يرتب MetaCrawler النتائج حسب درجة مطابقتها، وبحسب درجة المطابقة بالطريقة التالية:

أ- القيمة العظمى للمطابقة هي 1000 وتعني المطابقة الكاملة، والقيمة الدنيا هي 0 وتعني عدم مطابقة.

ب- بالاعتماد على قيم المطابقة التي أعادها كل محرك بحث للصفحات، يقوم MetaCrawler برد هذه القيم إلى المجال [0, 1000] مراعيا الوضع النسبي للنتائج فيما بينها. تمكن ملاحظة تغير درجة المطابقة بتغير محركات البحث المستخدمة، مثلا إذا قمنا بالبحث عن tourism في جميع محركات البحث التي يتيحها MetaCrawler، ثم أعدنا عملية البحث باستخدام بعض محركات البحث وليس جميعها، فسنجد أن تقييم المواقع المشتركة في نتائج البحثين قد تغير.

ج- الأرقام التي يعيدها هي تقديرات بدائية تعبر عن وجهة نظر MetaCrawler لذا من الطبيعي أن يجد المستخدم صفحات لا علاقة لها بالطلب الذي صاغه، خاصة إذا أخذنا بعين الاعتبار طريقة صياغته الطلبات (كلمات لازمة وممنوعة) وحقيقة أن بعض محركات البحث لا تستطيع التعامل مع هذا النوع من البحث، لذا تكون نتائجها تقريبية.

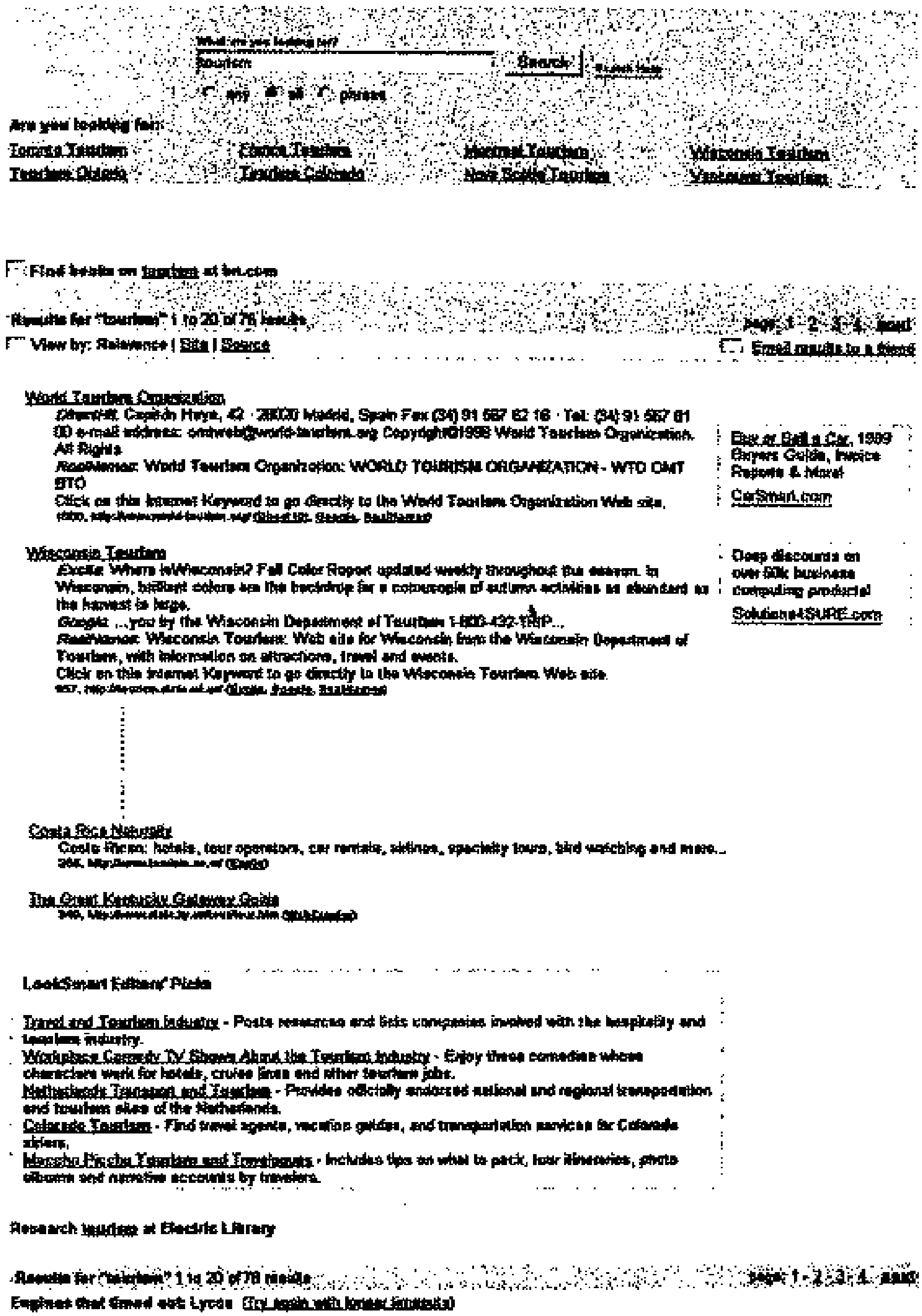
إضافة إلى ترتيب النتائج حسب درجات المطابقة، يستطيع MetaCrawler أن يرتبها حسب محرك البحث أو حسب Domain Name المعادة منه، ويمكن للمستخدم أن يحدد طريقة الترتيب التي يريدها.

2. يسمح بالفلترة حسب المنطقة الجغرافية.

3. يؤمن تصنيفا جيدا للمواضيع ضمن فئات بفضل مجلده.

4. إظهار النتائج: يظهر MetaCrawler نتائجه ضمن خمسة أقسام تظهر جميعها على الصفحة نفسها كما يبين ذلك الشكل (7–3). هذه الأقسام هي:

- أسفل مربع الحوار مباشرة، نجد مجموعة من طلبات البحث اشتقها MetaCrawler من طلب البحث، ويسأل المستخدم عنها بهدف مساعدته في صياغة البحث على نحو أفضل.
- بعد ذلك نجد كتابا مطابقا للطلب، ويعده عدد نتائج البحث المعادة من محركات البحث المختلفة، وهي هنا 76 نتيجة. يجب الانتباه إلى عدد النتائج المعادة، إذ يمكن عند تكرار طلب البحث أن نحصل على عدد مختلف. يعود السبب إلى أن MetaCrawler يعتمد على محركات البحث الأخرى من أجل الحصول على النتائج، وقد تكون محركات البحث هذه مزدحمة خلال فترة تنفيذ الطلب، مما يعطي عدد نتائج مختلفة. على يمين عدد النتائج، نجد ارتباطات تشعبية إلى باقي صفحات النتائج لأن الصفحة الواحدة لا تحوي أكثر من 20 نتيجة (يمكن تغيير عدد النتائج في الصفحة كما ذكرنا سابقا).



الشكل (7-3): نتائج البحث ضمن MetaCrawler

- بعد ذلك نجد مجموعة من الارتباطات التي، باختيار أحدها، نحدد طريقة ترتيب نتائج البحث المعادة. ونلاحظ أن طريقة ترتيب النتائج في الشكل (7–3) هي حسب المطابقة.
- يلي ذلك قسم عرض النتائج، حيث يعرض MetaCrawler كل نتيجة معادة كما يلي:
  - ارتباط تشعبي إلى الصفحة النتيجة، ثم شرحاً عن هذه الصفحة كما جاءه من محركات البحث. ففي الشكل (7–3)، نجد أن النتيجة الأولى قد أعادتها محركات البحث Direct Hit وGoogle وRealNames، أما الشروح، فالشرح الأول أعاده DirectHit والشرح الثاني أعاده RealNames، بينما لم يُعِد Googles أي شرح.
  - في آخر سطر من النتيجة، نجد أولاً درجة المطابقة التي أعطاها MetaCrawler للنتيجة، ثم نجد URL هذه النتيجة، ثم مجموعة المحركات التي وجدت هذه النتيجة.
  - في آخر صفحة النتائج نجد مجموعة من النتائج المعادة لنتيجة البحث ضمن مجلد LookSmart. وفي النهاية نجد مجموعة المحركات التي لم تتمكن من إعادة نتائجها خلال الوقت المحدد، وفي الشكل نجد أن Lycos لم يتمكن من إعادة نتائجه.

ج. البحث حسب الموصفات

يدعم MetaCrawler البحث حسب الموصفات ضمن مجموعة من أدواته مثل الصفحات الصفر، وأدوات البحث عن الأشخاص، وغيرها من الأدوات التي نجدها ضمن الجزء 6 (Resources).

7–2–3 إضافة موقع إلى محرك البحث MetaCrawler

لا يمكن إضافة موقع إلى MetaCrawler لأنه لا يملك قاعدة بيانات. كي يستطيع المستخدم رؤية صفحته (أو موقعه) بين نتائج MetaCrawler، يجب أن يُسجلها ضمن قاعدة بيانات أحد محركات

البحث المشهورة والتي يستخدمها MetaCrawler. لا يستطيع MetaCrawler الوصول إلى قواعد بيانات محركات البحث التي يستخدمها من أجل تسجيل صفحات المستخدم، لذا يجب على المستخدم أن يستخدم إحدى أدوات التسجيل (Register-It!!) من أجل تسجيل موقعه ضمن قواعد بيانات محركات البحث، من هذه الأدوات الأداة Register-It! التي عنوانها:

http://www.register-it.com/

والتي تستطيع تسجيل الموقع ضمن 400 محرك بحث.

7-3 محرك البحث البيني The Big Hub

WWW.theBigHub.com

يعتبر محرك البحث البيني theBigHub (وكان يسمى سابقا the Internet Sleuth) من أشهر وأسرع محركات البحث البينية، ويستطيع معالجة الطلبات التي تصاغ بلغة طبيعية.

7-3-1 واجهة theBigHub

يظهر الشكل (7-4) صفحة theBigHub على الوب، والتي يمكن الوصول إليها باستخدام العنوان: http:\\www. theBigHub.com. نجد فيما يلي الإمكانات المختلفة لـ theBigHub:

الجزء [1]: شعار theBigHub.

الجزء [2]: يسمح للمستخدم بالقيام بعمليات بحث.

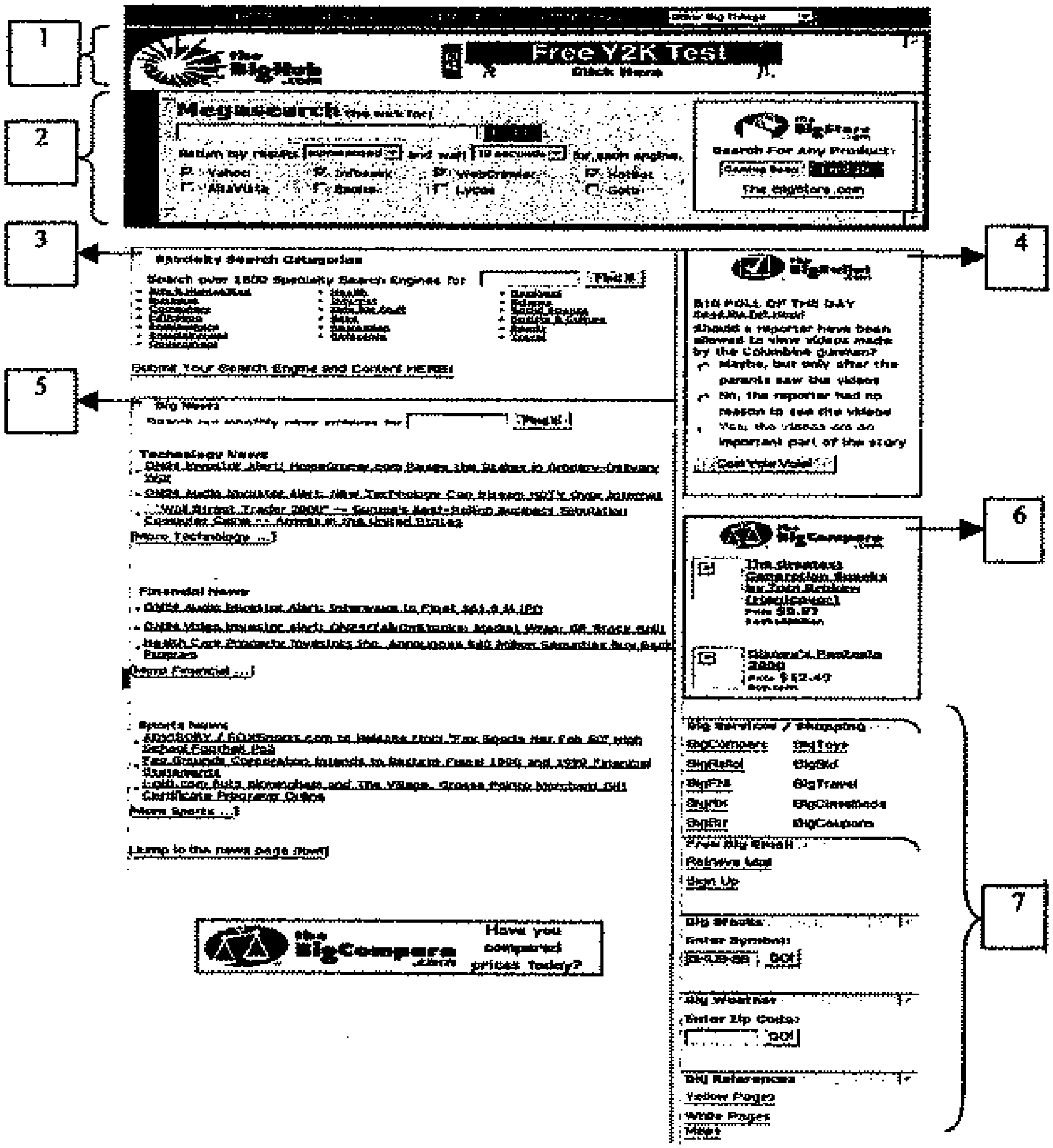
الجزء [3]: مجموعة من فئات البحث المخصصة.

الجزء [4]: أداة تسمح بالتصويت على مواضيع مطروحة.

الجزء [5]: أخبار.

الجزء [6]: أداة من أدوات theBigHub، سنتحدث عنها لاحقا.

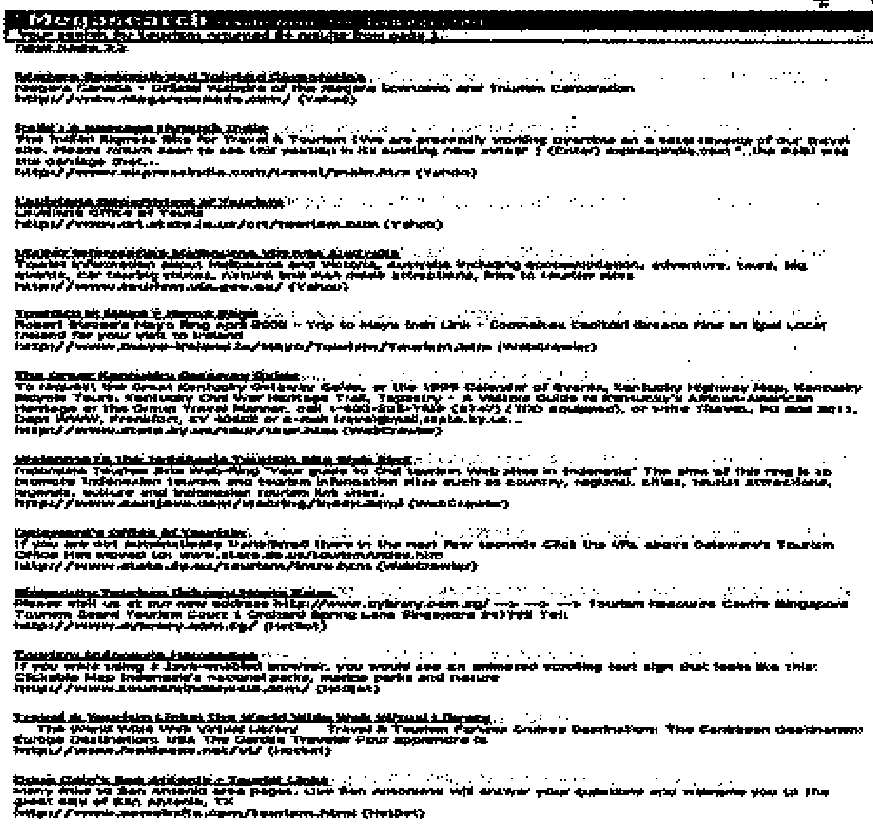
الجزء 7: مجموعة من أدوات وخدمات theBigHub.



الشكل (4-7): صفحة theBigHub.

7-3-2 البحث عن المعلومات في theBigHub

يتيح theBigHub إمكانية صياغة طلبات البحث من خلال الجزء [2] من صفحته الرئيسية. يصاغ طلب البحث بلغة طبيعية، ولا نجد أي إمكانات فلترة على النتائج. فكل ما يستطيع المستخدم تحديده هو المحركات التي سيبحث ضمنها والمهلة الزمنية التي يعطيها لكل محرك بحث، وطريقة عرض النتائج التي ترد من مجموعة محركات البحث المختارة، كما يظهر ذلك الشكل (7-5).



الشكل (7-5): صفحة النتائج في theBigHub

إن محركات البحث المختارة في مثالنا هنا هي Yahoo!, HotBot, WebCrawler, InfoSeek، ونجد أن النتائج تترتب حسب محركات البحث، ففي بداية صفحة النتائج، نجد مجموعة النتائج من Yahoo!، ثم النتائج من WebCrawler، ثم من HotBot، أما لرؤية نتائج InfoSeek فيجب الانتقال إلى الصفحة التالية من النتائج.

7-3-3 أدوات المعلومات الأخرى

The Big Compare: هي أول أداة تقدم خدمات مقارنة بين مواقع البيع على الوب. تفيد هذه الأداة في البحث عن سلعة ما (أو مادة أو غرض) وإيجاد المواقع التي تحويها، وعرضها للمستخدم بعد إجراء مقارنات عليها بحيث تحدد المواقع ذات الأسعار الأرخص والمواصفات الأجود.

7-4 محركات البحث العربية

تعتبر محركات البحث العربية في بداياتها، ومازالت تفتقر إلى إمكانية منافسة محركات البحث العالمية، كما أن قواعد بياناتها صغيرة جداً. الميزة الأساسية لهذه المحركات هي استخدام اللغة العربية سواء في واجهة الاستخدام أو في عمليات البحث والتصنيف.

اخترنا من بين محركات البحث العربية القليلة محرك البحث أين/نظرة.

7-5 محرك البحث العربي أين/نظرة

Ayna? أين؟
Your net guide دليلك على الإنترنت

WWW.Ayna.com

هو محرك البحث العربي الأول على الوب. أنشئ في عام 1997، ولا يزال يطمح إلى تقديم المزيد من الخدمات باللغة العربية بهدف تجميع جميع مستخدمي الوب الناطقين باللغة العربية في مكان واحد على الشبكة. يسعى القائمون على أين كي يصبح محركهم في العالم العربي مثل Yahoo! في العالم الغربي.

يستخدم أين محرك بحث يقبل طلبات باللغة العربية هو نظرة. وهو محرك بحث يطبق خوارزميات خاصة باللغة العربية من أجل تحليل مفردات النص العربي، واستخراج الكلمات المفتاحية منه.

7-5-1 واجهة أين

يظهر الشكل (7-6) صفحة أين على الوب، والتي يمكن الوصول إليها باستخدام العنوان: http:\\www.Ayna.com. نلاحظ على هذه الواجهة وجود مجلد خاص بـ أين، يمكن من خلاله التصفح ضمن فئاته المختلفة. وتمكن صياغة طلبات بحث ضمن حقل الحوار في أعلى الصفحة بأن يحدد المستخدم مجموعة من الكلمات المفتاحية التي تعبر عن طلبه.



الشكل (7-6): صفحة أين

الفصل الثامن

مقارنة بين محركات البحث

في هذا الفصل، سنقارن بين محركات البحث التي عرضناها آنفاً، ونبين الميزات التي يتفوق فيها أحد المحركات على غيره. كما سنورد مقارنات بالنسبة لعوامل أخرى، وقياسات لشعبية هذه المحركات عند المستخدمين. لا شكّ في أن قياس بعض العوامل هو أمر معقد، لذلك فقد استعنا عند الحاجة بنتائج بعض المواقع المتخصصة بمقارنة محركات البحث على الوب من أجل الحصول على بعض المعلومات الإضافية. هذه المواقع هي:

Search Engine Watch

http://www.searchenginewatch.com

Search Engine Guide

http://www.searchengineguide.org

وننصح جميع المستخدمين بزيارة هذين الموقعين، فهما يحتويان معلوماتٍ مفيدةً جداً عن محركات البحث. نختم هذا القسم بنصائح قد تفيد المستخدم عند التصفح والبحث في الوب.

8-1 البحث حسب المحتوى

يبين الجدول (8-1) مقارنة مباشرة وشاملة بين أشهر محركات البحث التي اخترناها وعرضناها سابقا، وذلك حسب المعايير التي حددناها.

	Yahoo!	Hot Bot	Alta Vista	Info Seek	Web Crawler	Lycos	Excite
تحسس للأحرف	✓	✓	✓	✓	✗	✗	✗
جمل	✓	✓	✓	✓	✓	✓	✓
كلمات لازمة	✓	✓	✓	✓	✓	✓	✓
كلمات ممنوعة	✓	✓	✓	✓	✓	✓	✓
رموز عامة	✓	✓	✓	✗	✗	✗	✗
ترتيب النتائج	✓	✓	✓	✓	?	?	✓
بحث بولياني	✗	✗	✓	✗	✓	✓	✓
بحث متقارب	✗	✗	✓	✗	✗	✓	✗
الفترة	✓	✓	✓	✓	✗	✓	✓
تصنيف مواضيع	✓	✓	✓	✓	✓	✓	✓
تغذية راجعة	✗	✓	✗	✓	✗	✗	✗
قاعدة البيانات	0.5[1]	110[2]	150	75	50	100	50

الجدول (8-1): مقارنة محركات البحث حسب المعايير المحددة

[1] نصف مليون موقع

[2] الوحدة لهذا المحرك والمحركات التي تليه هي مليون صفحة

8-1-1 التحسس للأحرف الصغيرة والكبيرة

تتحسس بعض محركات البحث لحالة الأحرف المستخدمة في الطلب، ويعود السبب في ذلك إلى رغبة محركات البحث هذه في إعطاء المستخدم القدرة على تضييق أو توسيع مجال البحث. فكما رأينا، تتحسس محركات البحث هذه لحالة الأحرف بنفس الطريقة، وهي "مراعاة حالة أحرف الطلب في حالة وجود حرف كبير"، مما يعني السماح للمستخدم بصياغة طلبات لا يراعي محرك البحث حالة الأحرف فيها (جميع الأحرف صغيرة)، أو صياغة طلبات نطلب فيها البحث عن كلمة محددة من حيث طريقة كتابتها (استخدام حرف كبير) كما في حالة البحث عن أسماء علم، أو عن اختصارات (IRS, SYR, USA).

8-1-2 الجمل

تدعم جميع محركات البحث المعروضة آنفاً البحث عن جمل، وتتفق جميعها في طريقة البحث (استخدام "")، ويؤمن بعضها طرقاً أخرى مثل HotBot الذي يسمح للمستخدم بتحديد نوع الطلب الذي صاغه (جملة، طلب بولياني،...). يتميز AltaVista عن أقرانه في هذا المجال، فهو يسمح باستخدام مجموعة من الرموز (_ ,/ ,# ,$,%) من أجل الربط بين الكلمات وتكوين الجملة، إضافةً إلى استخدام "".

8-1-3 الكلمات اللازمة والممنوعة

تدعمها جميع محركات البحث المعروضة، وبعض محركات البحث تعتبرها كافية، فلا تدعم العوامل البوليانية.

8-1-4 الرموز العامة

تقدم بعض محركات البحث هذه الميزة، لكن يتفرد HotBot بالتوسع في استخدامها، فهو يسمح باستخدام (?, *) في أي مكان من الكلمة، بينما لا يدعم AltaVista وYahoo! إلا استخدام * فقط وفي آخر الكلمة حصرا.

8-1-5 ترتيب النتائج

لا ترتب جميع محركات البحث نتائجها، أما تلك التي تفعل، فهي تقوم بذلك بطريقة غير جيدة (من وجهة نظرنا طبعا). فكما رأينا في الجزء الأول من هذا الكتاب يجري ترتيب نتائج البحث حسب درجات مطابقتها المتناقصة لطلب المستخدم. تكمن المشكلة في قياس درجة المطابقة هذه، ومدى صحتها فيما لو تمكنا من قياسها، وهو أمر يتعلق بتابع المطابقة المستخدم. إن بعض توابع المطابقة المستخدمة لا تسمح دائما بقياس درجة المطابقة (مثل تابع المطابقة البوليـاني)، عندئذ تعمد محركات البحث إلى ترتيب النتائج حسب عوامل أخرى لا ترتبط بمطابقة طلب المستخدم ارتباطا واضحا (على الأقل من الناحية النظرية). إن المبدأ العام في ترتيب النتائج لجميع محركات البحث هو موقع الكلمات ضمن الصفحة، وتواتر ورود هذه الكلمات، وينحصر الاختلاف بين محركات البحث في أمور وتفاصيل صغيرة. لعل أفضل محركات البحث المروضة ترتيبا لنتائجه هو InfoSeek، فهو قادر على قياس درجة المطابقة بين الطلب والصفحة، ومن ثم ترتيب النتائج حسب درجة المطابقة المتناقصة. يشابهه في طريقة الترتيب Excite الذي يستطيع قياس درجة المطابقة أيضا.

نلاحظ في الجدول وجود إشارة استفهام في حالة كل من WebCrawler وLycos فيما يخص ترتيب النتائج، وتعني، وبرغم ترتيب هذين المحركين لنتائجهما حسب درجة مطابقتها لطلب المستخدم، أننا لم نستطع إيجاد معلومات عن طريقة قياس المطابقة أو العوامل المؤثرة فيها.

8-1-6 البحث البولياني

لا تدعم جميع محركات البحث هذا النوع من البحث، ولكن تلك التي تتمتع بهذه الميزة تتشابه في العوامل التي تدعمها وهي AND, OR, NOT، عدا AltaVista الذي لا يسمح باستخدام NOT بل AND NOT كعامل واحد هو (!).

8-1-7 البحث المقارب

تدعم قلة من محركات البحث هذا النوع من البحث، ومن بين محركات البحث المعروضة نجد أن AltaVista و Lycos هما الوحيدان اللذان يدعمان هذا النوع من البحث. يسمح AltaVista باستخدام العامل NEAR فقط، بينما يؤمن Lycos مجموعة كبيرة من العوامل.

8-1-8 الفلترة

يسمح معظم محركات البحث بالقيام بعمليات فلترة على النتائج، ونجد أكثر إمكانات الفلترة عند HotBot، ولكن يجب الانتباه إلى وجود إمكانات فلترة عند محركات بحث أخرى غير متوفرة في HotBot.

8-1-9 تصنيف المواضيع

يُعتبر !Yahoo الأفضل في هذا المجال، ولنا وقفة أخرى مع مجلدات محركات البحث لاحقاً.

8-1-10 التغذية الراجعة

تدعم قلة من محركات البحث هذه الميزة (HotBot, InfoSeek).

8–1–11 قاعدة البيانات

إن أضخم قاعدة بيانات لمحرك بحث على الإطلاق هي قاعدة بيانات AltaVista، والسباق على أشده في اتجاه زيادة حجم قواعد البيانات مع تحسين دقة النتائج وسرعة البحث. نلفت الانتباه إلى أن Yahoo! ليس محرك بحث تماماً، بل هو مجلد تصنيف للمواقع على الوب، لذا فالرقم الموجود في خانة Yahoo! هو عدد المواقع ضمن مجلده، وليس عدد الصفحات. تمكن للمستخدم معرفة حجم قاعدة بيانات محرك بحث يدعم استخدام الرموز العامة والكلمات اللازمة من خلال الطلب "*+"، لأن حجم قاعدة البيانات سيكون هو عدد الصفحات التي تنتج. إن الأرقام الموجودة في الجدول هي الأرقام الرسمية التي تُصرح عنها محركات البحث، وفي إحدى الإحصائيات، قُدِّر عدد الصفحات على الوب بـ 800 مليون صفحة (1999/12)، ويمكننا أن نتبين من ذلك أن AltaVista لا يفهرس نسبة عالية من صفحات الوب.

أخيراً، توجد محركات بحث تدعم صياغة الطلبات الطبيعية مثل AltaVista وWebCrawler. إن الاتجاه نحو استخدام اللغات الطبيعية هو أمر لا غنى عنه، لكن يجب تنبيه المستخدم إلى أن النتائج التي سيحصل عليها باستخدام اللغات الطبيعية قد تكون سيئةً وغيرَ مطابقة لما يريد، كما ننبه إلى ضرورة استخدام أسلوب بسيط وجمل سهلة وواضحة في صياغة الطلبات.

8–2 استخدام التصنيف بالمجلدات

يُبين الجدول (8–2) مقارنة بين محركات البحث حسب خصائص مجلداتها (فئات الموضوعات).

8–2–1 النمط

يدل على نمط الخدمة الأصلي، أهو مصنِّف أم محرك بحث[3]؟.

---

[3] توجد خدمة ثالثة هي خدمة الإجابة على الأسئلة، كما في *AskJeeves* (*http://www.askjeeves.com*).

8-2-2 المحررون

يمثل عدد القائمين على المجلد. وارتفاع عدد المحررين ليس أفضل بالضرورة، بسبب استخدام بعض المجلدات لتقنيات مساعدة، لكن العدد الكبير إشارة جيدة على أن المجلد مبني جيدا وأنه يتطور وينمو بنمو الوب.

8-2-3 الفئات

يمثل عدد فئات الموضوعات في كل مجلد.

8-2-4 الارتباط

يمثل عدد مداخل المجلد مقدرا بمليون ارتباط. يمكن أن يحوي المجلد مواقع بكاملها أو صفحات مستقلة (حسب المضمون). كما يمكن للموقع أن يدرج ضمن أكثر من فئة.

8-2-5 التاريخ

وهو التاريخ الذي جمعنا فيه هذه النتائج.

الخدمة	النمط	المحررون	الفئات	الارتباط	التاريخ
Yahoo	مصنف	100+	25,000	1.2+	11/99
LookSmart	مصنف	200	60,000	1	9/99
Go (InfoSeek)	محرك بحث	2,600	50,000	0.5	10/99
Open Directory	مصنف	15,400	153,000	0.95	9/99
AltaVista	محرك بحث	انظر LookSmart			
Excite	محرك بحث	انظر LookSmart			
HotBot	محرك بحث	انظر Open Directory			
Lycos	مجلد	انظر Open Directory			

الجدول (8-2): مقارنة بين محركات البحث حسب خصائص مجلداتها

نلاحظ من الجدول السابق أن AltaVista لا يملك مجلدا خاصا به، بل يستفيد من مجلد LookSmart لتقديم خدمة التصفح بين المواقع حسب مواضيعها ومثله Excite، أما Lycos وHotBot فهما يستفيدان من المجلد المفتوح كما أشرنا عند الحديث عنهما. ونذكر بأن AltaVista سيشترك في المجلد المفتوح قريبا.

8-3 تسجيل المواقع ضمن محركات البحث

يبين الجدول (8-3) مقارنة بين محركات البحث حسب طريقة تسجيل المواقع والصفحات فيها.

8-3-1 زمن التسجيل

نلاحظ أن AltaVista يحتاج إلى أقصر زمن من أجل ظهور الصفحات المسجلة ضمن فهرسه، ولكنه يحتاج إلى 3 أسابيع كي تنقل التغييرات إلى المواقع المرايا لـ AltaVista (Mirror Sites).

8-3-2 يدوي/آلي

طريقة عمل محرك البحث من أجل العثور على صفحات جديدة وفهرستها. نلاحظ أن Yahoo! هو الوحيد الذي يقوم بعمل يدوي، باعتبار أن عمله يقسم إلى قسمين يدوي وآلي. يملك Yahoo! روبوتا خاصا به من أجل البحث فقط عن مواقع جديدة على الوب (URL)، أما الفهرسة والتصنيف فيقوم بها أشخاص معنيون في Yahoo!، أي يدويا. يفيد استخدام روبوت (كنا قد سميناه عنكبوتا آليا في مواضع سابقة من الكتاب) في عملية صيانة المجلد والحفاظ عليه، لأن عملية التحقق من وجود ارتباطات تشعبية غير صالحة ضمن المجلد تصبح أسهل.

تقوم بقية محركات البحث بعملية تصفح آلي على الوب، وتجمع صفحات جديدة، وتفهرسها آليا. تطلب محركات البحث جميعا عند تسجيل موقع من قبل مستخدم عنوان الموقع فقط. يختلف

HotBot قليلا في هذا، فهو يعطي إمكانية قيام المستخدم بتسجيل مجموعة من صفحات موقعه من أجل اختصار الزمن، لذا فالعمل يمكن أن يكون يدويا.

8-3-3 خصائص وحدود

وهي الشروط التي يضعها محرك البحث عند تسجيل صفحة أو موقع. لا يضع Yahoo! أي شروط، بينما لم نتمكن من العثور على أي معلومات بهذا الخصوص عن AltaVista.

تسجيل موقع	Yahoo!	Hot Bot	Alta Vista	Info Seek	Web Crawler	Lycos	Excite
زمن التسجيل	يومان	21-60 يوما	يوم - 6 أسابيع	6 أسابيع	أسبوعان	2-3 أسابيع	أسبوعان
يدوي/آلي	يدوي	آلي/يدوي	آلي	آلي	آلي	آلي	آلي
خصائص وحدود	/	أصغر من 50 صفحة من نفس الموقع في اليوم			25 صفحة من الموقع، حجم الصفحة < 1 ميغا بايت	عدم استخدام بعض الرموز ضمن URL الموقع	
عقوبات الغش	غير ممكنة	خفض درجة مطابقة صفحات الموقع	حذف الموقع	حذف الموقع	حذف الموقع		
HTML Meta Tags	/	75 حرفا للكلمات المفتاحية، و150 حرفا للشرح	✓	1000 حرف كلمات مفتاحية، 200 حرف شرح		×	إضافة شرح فقط
منع الفهرسة	/	Robots.txt HTML Meta Tags	Robots.txt HTML Meta Tags	Robots.txt	Robots.txt		Robots.txt

الجدول (8-3): مقارنة بين محركات البحث حسب طريقة تسجيل المواقع والصفحات

8-3-4 عقوبات على محاولات الغش

وهي العقوبة التي يفرضها محرك البحث على المواقع التي يكتشف قيامها بمحاولات غش. معظم محركات البحث تحذف هذه المواقع من فهارسها، أما HotBot فيكتفي بإنقاص درجة مطابقة الصفحات التي في هذه المواقع. لا يمكن القيام بأي محاولة غش مع Yahoo! لأن العمل يدوي. لم نتمكن من معرفة تصرف WebCrawler تجاه محاولات الغش.

8-3-5 HTML Meta Tags

وهي مؤشرات HTML تسمح للمستخدم بتحديد كلمات مفتاحية ووضع شرح عن موقعه أو صفحته. لا يسمح lycos بهذا الأمر، فيما يسمح بها كل من InfoSeek, HotBot, AltaVista بشروط معينة، ولم نتمكن من معرفة تصرف WebCrawler تجاهها.

8-3-6 منع الفهرسة

وهي الطريقة التي يمكن اتباعها من أجل منع فهرسة صفحة من صفحات الموقع. يمكن اتباع أحد أسلوبين: Robots.txt أو استخدام مؤشرات HTML الخاصة.

8-4 شعبية محركات البحث

يبين الجدول (8-4) إحصائية عن شعبية محركات البحث مقيسة بنسبة المستخدمين الذين يستخدمون محرك البحث. نجد Yahoo! في الصدارة بلا منازع، ويليه AltaVista. يلاحظ أن مجموع النسب لا يساوي 100، والسبب يكمن في أن مستخدمي محركات البحث الذين شملتهم الإحصائية يستخدمون أكثر من محرك بحث واحد.

حتى نوضح الصورة العامة على نحو أفضل، نورد الجدول (8-5) الذي يعطي فكرة عن عدد زوار مواقع محركات البحث، والعدد الوسطي للصفحات التي يتصفحها كل زائر شهريا (6/1998).

محرك البحث	الشعبية
Yahoo!	93.4%
AltaVista	81.4%
InfoSeek	70.2%
Excite	69.9%
Lycos	68.2%
HotBot	46.2%
WebCrawler	45.6%

الجدول (8-4): شعبية محركات البحث

الخدمة	عدد الزوار (مليون)	عدد الصفحات
Yahoo!	26.5	162
Excite	13	40
InfoSeek	10.7	30
Lycos	9.4	23
AltaVista	6.7	20
WebCrawler	5.2	19
HotBot	4.3	19

الجدول (8-5): عدد زوار محركات البحث والعدد الوسطي للصفحات التي يتصفحها كل زائر شهريا

5-8 أهمية محركات البحث

لعل أحد الأسئلة التي تخطر على الأذهان هو "هل محركات البحث هي على ذلك القدر من الأهمية والفائدة لمن يتصفح الوب ؟". سنحاول الإجابة على هذا السؤال من خلال الجدول والمخطط التاليين:

يعطي الجدول (8-6) فكرة عن النسبة المئوية للمستخدمين والوقت الذي يقضونه في البحث باستخدام محركات البحث يوميا.

إذا اعتبرنا أن زمن القيام ببحث جدي وفعلي هو أكبر من 5 دقائق، فإننا نجد أن نسبة لا بأس بها من المستخدمين تقوم بعمليات بحث فعلية (أكثر من 60٪). لكن في نفس الوقت، نجد أن نسبة المستخدمين الذين يستخدمون محركات البحث كنقطة انطلاق للتصفح في الوب لا بأس بها. ونقصد بذلك الأشخاص الذين يستخدمون محرك البحث (أقل من 5 دقائق) من أجل القيام ببحث سريع عن موقع لا يعرفون عنوانه.

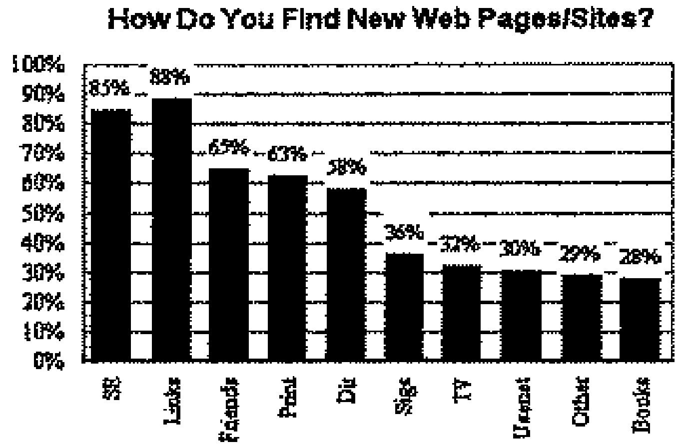
نسبة مئوية	الزمن
29.3%	<5 min.
35.2%	5-10 min.
17.2%	10-20 min.
6.0%	20-30 min.
1.9 %	30 – 60 min.
4.8%	>60 min.
3.5%	Don't know
1.6%	Not Applicable

الجدول (8-6): استخدام محركات البحث

نتبين من الشكل (8-1) (أخذ في 1998/12)، أهم مصادر المعلومات التي يستعملها المستخدمون من أجل العثور على عناوين المواقع.

نلاحظ أن مجموع هذه النسب لا يساوي المائة لأن المستخدمين الذين شملهم الإحصاء بإمكانهم اختيار أكثر من طريقة للحصول على عناوين المواقع.

إن أهم الطرق المستخدمة هي محركات البحث (85٪)، والبحث بالتصفح باتباع الارتباطات التشعبية (88٪). فمحركات البحث إذن لا تزال على درجة عالية من الأهمية لمستخدمي الوب من أجل العثور على عناوين المواقع.



الشكل (8-1): نسب الوصول إلى مواقع/صفحات وب جديدة

بقية الطرق هي:

Friends: الأصدقاء، **Print**: المطبوعات من مجلات وصحف، **Dir**: مصنفات الإنترنت مثل Yahoo!، وهنا يعتقد القائمون على الإحصاء أن المعلومات غير دقيقة بسبب عدم تمييز الكثير من المستخدمين للفرق بين المجلد ومحرك البحث، **Sigs**: التواقيع في نهاية رسائل E-mail، **TV**: إعلانات التلفاز، **Usenet**: مجموعات usenet الأخبارية، **Other**: موارد أخرى، **Books**: الكتب.

6-8 أزمنة استخدام محركات البحث

تعطي المقارنة بين أزمنة استخدام محركات البحث فكرة عن نظرة المستخدمين إلى محرك البحث، هل هو مكان لصياغة الطلبات ثم الانتقال إلى المواقع التي يعطيها، أم هو محطة يقصدها المستخدمون لذاتها وللخدمات التي تقدمها ؟. يبين الجدول (8-7) الأزمنة التي يقضيها المستخدم وسطيا عند استخدام كل محرك بحث خلال شهر (1999/9).

محرك البحث	الزمن
yahoo.com	1:04:33
Excite.com	0:29:53
altavista.com	0:08:47
Lycos.com	0:08:16
webcrawler.com	0:06:46
hotbot.com	0:05:04

الجدول (8-7) : الأزمنة التي يقضيها المستخدمون في استخدام محركات البحث

يدل الجدول السابق على أن المستخدمين لا يعتبرون Yahoo! محرك بحث فقط، بل هو مكان يقصده معظم مستخدمي الإنترنت.

الفصل التاسع

خلاصة وتوصيات

9-1 بعض النصائح للباحثين عن المعلومات في الإنترنت

كما وجدنا في هذا الجزء، هناك مجموعة ضخمة من أدوات البحث في الإنترنت. ومن الصعب المقارنة بين هذه الطرق لأن كلاً منها لها إيجابياتها وسلبياتها، والحقيقة أنه بسبب الطبيعة الشبكية للمعلومات فإن هذه الأدوات تخدمنا أكثر عندما تعمل بعضها مع بعض.

لذلك فإنه من أجل عملية بحث معينة قد يصعب اختيار الأداة المناسبة، وبفرض أننا اخترنا هذه الأداة فهل هي كافية للحصول على البحث المطلوب ؟.

من تجربتي الشخصية أستطيع تقديم النصائح التالية:

* البحث عن المعلومات في الإنترنت يتطلب الكثير من الصبر والهدوء لأن هناك الكثير من التجربة/الخطأ (Try and Error) وبالتالي الكثير من الإحباط.
* من خلال البحث المطلوب، اختر الأداة المناسبة ولا تتردد باختيار أدوات أخرى في عند فشل البحث باستخدام الأداة التي اخترتها مبدئياً.

* تعلم أكثر ما تستطيع عن أداة البحث المستعملة وخصوصاً ما يتعلق بقواعد لغة كتابة الطلبات وكيفية الفلترة، وأكثر من التجول لأن ذلك يعطيك فكرة ولو بسيطة عن كيفية ترتيب فضاء المعلومات.
* وأخيراً حاول أن تضع نفسك مكان مصمم أداة البحث وتمثل منطقه الخاص وطريقته في الفهرسة والتبويب فإن ذلك يمكن أن يكون مفيداً جداً.

9-2 بعض التصورات المستقبلية للبحث عن المعلومات في الإنترنت

قد لا يكون رهان المستقبل على إضافة أدوات جديدة بقدر ما سيكون على تطوير الأدوات الموجودة بهدف الوصول إلى نتائج أفضل وخصوصاً باتجاه تحسين الواجهات التخاطبية لمحركات البحث، وتطوير نماذج البحث المعتمدة حالياً على النموذج البولياني إلى النماذج المعتمدة على معالجة اللغات الطبيعية وقواعد المعارف، وكذلك نحو ترتيب النتائج حسب درجة المواءمة (Relevance Ranking) ونحو التغذية الراجعة (Relevance Feedback).

يتطلب كلُّ ذلك إيجاد نماذج نظرية جديدة تستطيع معالجة المعارف وبناء الكتل (Clusters) ومكاملة التصفح على مستوى النموذج، كما يتطلب تسريع البحث أيضاً بالاستفادة من التقدم التكنولوجي في الاتصالات والمعالجة والتخزين.

ملحق

بعض العناوين المفيدة في البحث عن المعلومات في الإنترنت

يحتوي هذا الملحق عناوين URL أشهر مواقع محركات البحث على الوب. ونقسمها إلى ستة أقسام حسب استخدام محركات البحث هذه (يمكن ورود محرك بحث ضمن أكثر من قسم). هذه الأقسام هي:

1. محركات البحث العامة

من بين مئات محركات البحث المتوفرة على الإنترنت، اخترنا فيما يلي أشهرها:

محرك البحث About	About: http://www.about.com/
محرك بحث بعيد عن المألوف، يُصاغ طلب البحث فيه بنقر كلمات البحث بدلاً من كتابتها، تجب قراءة التعليمات قبل استخدامه	Aeiwi: http://www.aeiwi.com/
محرك البحث العربي الإدريسي	Al-Idrisi: http://www.idrisi.com/
محرك بحث أكاديمي	AliWeb: http://www.aliweb.com/

AltaVista: http://www.altavista.com/	محرك البحث AltaVista
AOL Search: http://search.aol.com/	موقع أمريكا أون لاين
Ask Jeeves: http://www.askjeeves.com/	أداة بحثٍ الطلبُ فيها هو سؤالٌ، والنتيجة هي مجموعة من المواقع التي تجيب على السؤال
Ayna: http://www.ayna.com/	محرك البحث العربي أين
BigBook: http://www.bigbook.com/	محرك البحث Big Book
BigFoot: http://www.bigfoot.com/	محرك البحث Big Foot
Britannica Internet Guide: http://www.britannica.com/	مجلد لمواقع الوب معدٌّ من قبل ناشري الموسوعة البريطانية Britannica
Cyper411: http://www.cyber411.com/	محرك البحث Cyber 411
Diabolos: http://www.diabolos.com/	محرك بحث أوروبي
Direct Hit: http://www.directhit.com/	محرك البحث Direct Hit
EuroSeek: http://www.euroseek.net/	محرك بحث أوروبي، بمختلف اللغات الأوروبية
Excite: http://www.excite.com/	محرك البحث Excite
EZResult: http://www.ezresult.com/	محرك البحث EZResult
FAST Search: http://www.alltheweb.com/	محرك البحث All the Web

GOD (Global Online Directory): http://www.god.co.uk/	مجلد مقره بريطانيا
Google: http://www.google.com/	محرك البحث Google
GoTo: http://www.goto.com/	محرك بحث يبيع نتائجه، حيث بإمكان الشركات أن تدفع نقودا لتحصل على ترتيب أعلى ضمن نتائج البحث
HotBot: http://www.hotbot.com/	محرك البحث Hotbot
InfoSeek: http://www.infoseek.com/	محرك البحث InfoSeek
InfoSpace: http://www.infospace.com/	محرك البحث Info Space
Inktomi: http://www.inktomi.com/	محرك البحث Inktomi
LookSmart: http://www.looksmart.com/	مجلد يدوي
LookSmart: http://www.looksmart.com/	محرك البحث Look Smart
Lycos: http://www.lycos.com/	محرك البحث Lycos
Magellan: http://www.magellan.com/	محرك البحث Magellan
MSN Search: http://search.msn.com/	خدمة البحث الخاصة بشبكة Microsoft
Northern Light: http://www.northernlight.com/	محرك البحث Northern Light
Open Directory: http://dmoz.org/	أضخم مجلد يدوي

محرك البحث Open Text	OpenText: http://www.opentext.com/
محرك البحث Planet Search	PlanetSearch: http://www.planetsearcht.com/
محرك البحث RealNames	RealNames: http://www.realnames.com/
مجلد من المواقع. بطيء غالباً	SearchKing: http://www.searchking.com/
مجلد يدوي	Snap: http://www.snap.com
محرك البحث Voila	Voila: http://www.voila.com/
محرك البحث WebCrawler	WebCrawler: http://www.webcrawler.com/
محرك بحث يعطي نتائج واردة من استراليا ونيوزيلاندة فقط	WebWombat: http://www.webwombat.com.au/
محرك البحث (أو مصنف) Yahoo!	Yahoo!: http://www.yahoo.com/

2. محركات البحث المخصصة للأطفال

نجد فيما يلي مجموعة من المحركات والمجلدات تضمن عدم ظهور أمور مؤذية للأطفال كنتيجة لعمليات البحث.

مجموعة ارتباطات إلى مواقع آمنة من شبكة America On Line	AOL NetFind Kids Only: http://www.aol.com/netfind/kids/
موقع للإجابة على الأسئلة	Ask Jeeves For Kids: http://www.ajkids.com/

يحوي 5000 موقع على الوب	KidsClick!: http://sunsite.berkeley.edu/KidsClick!/
مخصص من Yahoo! للأطفال من 7 إلى 12 سنة	Yahooligans: http://www.yahooligans.com/

3. محركات البحث البينية

يوجد العديد من المحركات البينية ، نذكر أشهرها:

محرك البحث البيني Total Search	C4 TotalSearch Technology: http://www.c4.com/
محرك البحث البيني Dog Pile	DogPile: http://www.dogpile.com/
محرك البحث البيني Inference Find	Inference Find: http://www.infind.com/
محرك البحث البيني Mamma	Mamma: http://www. mamma.com/
محرك البحث البيني MetaCrawler	MetaCrawler: http://www.metacrawler.com/
محرك البحث البيني ProFusion	ProFusion: http://www.profusion.com/
محرك البحث البيني SavvySearch	SavvySearch: http://www.savvysearch.com/
محرك البحث البيني Starting Point	StartingPoint: http://www.startingpoint.com/
محرك البحث البيني TheBigHub	TheBigHub: http://www.thebighub.com/

يوجد أيضاً نوع خاص من محركات البحث البينية يسمى: All-In-One. وهو بخلاف محركات البحث البينية العادية، لا يُرسل طلب البحث إلى مجموعة من محركات البحث، بل يعرض مجموعة من محركات البحث، ويسمح للمستخدم بتحديد محرك البحث الذي سيُرسَل إليه الطلب. نذكر من هذه المحركات ما يلي:

All4one Search Machine:
http://www.all4one.com

Instantseek:

http://instantseek.8m.com/

One Page MultiSearch Engines:
http://www.bjorgul.com

OneSeek:

http://www.oneseek.com/

Proteus:

http://www.thrall.org/proteus.html

PureSearch:

http://www.puresearch.com/

Search Spaniel:
http://www.searchspaniel.com/

Search-It-All:

http://www..search-it-all.com/

Skworm:

http://www.skworm.com/

مواقع لمعلومات عن محركات البحث البينية:

Cnet Guide to Metasearchers: April 1999:

http://home.cnet.com/category/topic/0,10000,0-3817-7-276933,00.html

Metasearch engines: precision searching:

http://www.ijs.co.nz/snoozinf.htm

4. محركات البحث عن الوسائط المتعددة

محركات البحث التي نعرضها هنا تبحث فقط عن الوسائط المتعددة (صورة، فيديو، صوت، ...):

AltaVista Photo Finder:

http://image.altavista.com/cgi-bin/avncgi

Ditto:

http://www.ditto.com/

Lycos Pictures and Sounds:

http://www.lycos.com/picturethis/

Scour.Net:

http://scour.net/

MediaFind:

http://search.mp3.de/

MIDI Explorer:

http://www.musicrobot.com/

StreamSearch.com:

http://www.streamsearch.com/

كما توجد محركات بحث أخرى تبحث عن ملفات MP3 فقط، منها:

Lycos / FAST MP3 Search:

http://mp3.lycos.com/

MP3.com:

http://mp3.com/

MP3meta:

http://www.mp3meta.com/

2Look4:

http://www.2look4.com/

AudioGalaxy:

http://www.audiogalaxy.com/

Oth.net:

http://oth.net/

Audiofind:

http://www.audiofind.com/

MediaFind:

http://search.mp3.de/

MediaLeech

search: http://medialeech.m4d.com/

Arianna MP3:

http://mp3.iol.it/

Getsongs - MP3 Search Engines:

http://altern.org/getsongs/

Soundcrawler:

http://www.soundcrawler.com/

5. محركات البحث الإقليمية

وهي ليست جميعاً محركات بحث، فبعضها لا يعدو أن يكون مجلدات تفهرس مجموعة من المواقع في الشرق الأوسط. نختار منها ما يلي:

AME Info - Middle East Business Information: http://www.ameinfo.com/	يغطي أكثر من 200000 شركة من 13 بلداً في الشرق الأوسط
Arabist: http://www.arabist.com/	دليل إلى مجموعة من المواقع
Iran Index: http://www.iranindex.com	محرك بحث إيراني
Libanis: http://www.libanis.com/	مجلد من المواقع في لبنان

مجلد شبيه بـ Yahoo!	Middle East Internet Pages: http://www.middle-east-pages.com/
مجلد من مجموعة من الفئات، بحث بالتجول، أخبار وحالة الطقس، وغيرها...	Orientation Middle East: http://me.orientation.com/
مجلد خاص بالإمارات العربية المتحدة	Orientation United Arab Emirates: http://uae.orientation.com/
مجلد عن المواقع في سورية	Syria Gate: http://www.syriagate.com/

6. محركات البحث عن الأخبار

اخترنا فيما يلي أشهر محركات البحث التي تبحث في الأخبار:

محرك بحث جديد	1stHeadlines: http://www.1stheadlines.com/
أخبار من 300 مصدر أخبار في كندا	AltaVista Canada - Canadian News Index: http://www.altavistacanada.com/
محرك بحث عام في الأخبار	DejaNews: http://www.dejanews.com/
خدمة البحث عن الأخبار من Excite فقط	Excite NewsTracker: http://nt.excite.com
يجمع أخبارا من المواقع الموجودة في جنوب إفريقية	Fanagalo: http://www.fanagalo.co.za/
خدمة بحث عن الأخبار من HotBot فقط	HotBot News Search / NewsBot: http://www.newsbot.com/
يبحث في مقالات صادرة عن أكثر من 4000 نشرة إلكترونية	InfoJump: http://www.infojump.com/

News Index: http://www.newsindex.com/	يفهرس الأخبار من مئات المصادر المختلفة. يُحدّث معلوماته بمعدل مرة كلّ ساعة.
NewsHub: http://www.newshub.com/	محرك بحث عام عن الأخبار
NewsNow: http://www.newsnow.co.uk/	البحث عن أخبار في 150 مصدر مختلف لا يتجاوز تاريخ نشرها ثلاثين يوماً
NewsTrawler: http://www.newstrawler.com/	يسمح بإرسال طلبات بحث إلى أكثر من محرك بحث
Northern Light's Current News: http://www.northernlight.com/news.html	خدمة البحث عن الأخبار من Northern Light's فقط.
Paperboy: http://www.paperboy.de/	البحث عن أخبار من ألمانيا
TotalNews: http://www.totalnews.com	محرك بحث عام عن الأخبار

المراجع المستخدمة في الكتاب

* جميع مواقع الإنترنت المذكورة في الكتاب.

[Agosti 93] M. AGOSTI. *Hypertext and Information Retrieval.* Information Processing & Management, Vol. 29, N 3, pp. 283-286, 1993.

[Chevalet 92] Jean Pierre CHEVALET. *Un modèle logique de Recherche d'Informations appliqué au formalisme des Graphes Conceptuels, le prototype ELEN et son expérimentation sur un corpus de composants logiciels.* Thèse de l'université Joseph Fourier, Grenoble, FRANCE, 15 mai 1995.

[Croft 93] W. Bruce CROFT, Howard TRURTLE. *Retrieval Strategies for Hypertext.* Information Processing & Management, Vol. 29, N 3, pp. 313-324, 1993.

[Dunlop 93] M. D. DUNLOP, C. J. RIJSBERGEN. *Hypermedia and Free Text Retrieval.* Information Processing & Management, Vol. 29, N 3, pp. 287-298, 1993.

[Goldfarb 90] Charles F. GOLDFARB. *The SGML HandBook.* Calrendon Press, Oxford, 1990.

[Halasz 88] Franck G. HALASAZ. *Refelections on Notecards: Seven Issues for the next generation of Hypermedia Systems.* Communications of the ACM, Vol. 37, N 2, pp. 836-852, 1988.

[Johnson 96] Scott JOHNSON. *Electronic Publishing Construction Kit.* Copyright © 1996 by John Willey & Sons, Inc. USA.

[Kheirbek 95a] Ammar KHEIRBEK, Yves CHIARAMELLEA. *Integrating Hyprmedia and Information Retrieval with Conceptual Graphs.* HIM'95, Konstanz, GERMANY, April 1995.

[Kheirbek 95b] Ammar KHEIRBEK. *Modèle d'intégration d'un Système de Recherche d'Information et d'un Système Hypermédia basé sur le formalisme des Graphes Conceptuels.* Thèse de l'université

Joseph Fourier, Grenoble, FRANCE, 29 mai 1995.

[Kheirbek 96] Ammar KHEIRBEK, Yves CHIARAMELLA. *Integrated model for Hypermedia and Information Retrieval*. In Information Retrieval and Hypertext, M. Agosti and A. Smeaton (Ed), Kluwer Academic (Publ.), pp. 139-178, 1996.

[Kheirbek 97] Ammar KHEIRBEK, Yves CHIARAMELLEA. *Browsing and Querying: two complementary approaches for Multimedia Information Retrieval*. HIM'97, Dortumond, GERMANY, September 1997.

[Li 92] Z. LI, H. DAVIS, W. HALL. *Hypermedia Links and Information Retrieval*. 14th Proceedings of the Information Retrieval Colloquim, Lancaster, pp. 169-180. 1992.

[Raggett 98] Dave Raggett, Arnaud Le Hors, Ian Jacobs. *HTML 4.0 Specification, W3C Recommendation.* http://www.w3.org/TR/REC-html40

[Rijebergen 86] C. J. RIJSBERGEN. *A new Theoritical Framework for Information Retrieval*. Proceedings of the ACM Conference on Research & Development in Information Retrieval, Pisa, ITALY, pp. 194-200, Sep. 1986.

[Sowa 84] Hohn F. SOWA. *Conceptual Structures: Information Processing in Mind and Machine*. Addison-Wesley publishing company, 1984.

[Weide90] Th. V. D. WEIDE, P. D. BRUZA. *Two-Level Hypermedia: An improved architecture for Hypertext*. Proceedings of the Database and Expert System Applications, Dexa'90, Springer Verlag, Vienna, AUTRICHE, Sep. 1990.

عناوين صدرت في سلسلة الرضا للمعلومات

اسم الكتاب	المؤلف	تاريخ النشر
١- بيئة النوافذ WINDOWS 3.11	م. أحمد شريك	١٩٩٤
٢- مبادئ الصيانة والشبكات	م. عبد الله أحمد	١٩٩٤
٣- معالجة النصوص MS WORD 6.0	د. هيثم البيطار	١٩٩٥
٤- ادخل إلى عالم WINDOWS 95	م. مهيب النقري	١٩٩٦
٥- قواعد البيانات MS ACCESS	زياد كمرجي - بيداء الزير	١٩٩٧
٦- توابع وماكروات في MS EXCEL 97	أ. زياد كمرجي	١٩٩٧
٧- مرجع تعليمي شامل لبرنامج معالجة النصوص MS WORD 97	د. هيثم البيطار	١٩٩٧
٨- مرجع تعليمي شامل في MS EXCEL 97	أ. زياد كمرجي	١٩٩٧
٩- مرجع تعليمي شامل في صيانة الحواسب الشخصية	م. عبد الله أحمد	١٩٩٨
١٠- مرجع تعليمي في برنامج الرسم والتصميم الهندسي AUTOCAD 14	م. احسان مردود	١٩٩٨
١١- المرجع التدريبي الشامل لـ WINDOWS 98	م. إياد زوكار	١٩٩٨
١٢- ادخل إلى عالم WINDOWS 98	م. مهيب فواز النقري	١٩٩٨
١٣- الإنترنيت وإنترانيت وتصميم المواقع	م. عبد الله أحمد	١٩٩٨
١٤- تكنولوجيا المعلومات على أعتاب القرن الحادي والعشرين	هاني شحادة الخوري	١٩٩٨

١٥-الإدارة الاستراتيجية للشركات والمؤسسات	د.يونس حيدر	١٩٩٩
١٦-نظام الـ ISO 9004-1	م.محمد حسن -م.بسام عزام	١٩٩٩
١٧-القائد المفكر حافظ الأسد		
والمشروع التنموي الحضاري	د.رياض عواد-أ.هاني الخوري	١٩٩٩
١٨- فن إدارة البشر	د. محمد مرعي مرعي	١٩٩٩
١٩- المرجع الشامل لتعليمات		
برنامج AUTOCAD	م. احسان المردود -م. وهبي معاد	١٩٩٩
٢٠- الدعاية والتسويق ومعاملة الزبائن	م. حنا بللوز	١٩٩٩
٢١- المعلومياء (المعلوماتية)		
ظروفها وآثارها الاقتصادية - الاجتماعية	د. معن النقري	١٩٩٩
٢٢- المرجع الشامل لبرنامج		
3D STUDIO MAX - الجزء الأول	م. جورج عطا الله بركات	١٩٩٩
٢٣- دليل الجودة في المؤسسات والشركات	د. طلال عبود-أ.ماهر العجي	١٩٩٩
٢٤-المرجع المفيد في علم شبكات الحواسيب		
	د.معتصم شفا عمري	١٩٩٩.
٢٥- ادخل إلى عالم ORACLE 8	م. مهيب النقري	١٩٩٩
٢٦- أسس إدارة الموارد البشرية	د. محمد مرعي مرعي	١٩٩٩
٢٧- تعلم برنامج إدارة قواعد البيانات	أ. زياد كمرجي - م. مهيب النقري	١٩٩٩
٢٨- الدليل الشامل لأساسيات		
الحاسوب والمعلوماتية	م. عبد الله أحمد	١٩٩٩
٢٩- الكذبات العشر للعولمة	د. عدنان سليمان	١٩٩٩
٣٠- بعض مسائل الاقتصاد اللاسياسي	د. مطانيوس حبيب	١٩٩٩
٣١- دليل إعادة تنظيم المؤسسات	د. محمد مرعي مرعي	١٩٩٩

٣٢- الدراسات التسويقية ونظم معلومات التسويق	د. طلال عبود - د. حسين علي	١٩٩٩
٣٣- مدخل إلى المعلوماتية الطبية	م. جورج بركات - أ. هاني الخوري	١٩٩٩
٣٤- الدعاية والتسويق وفن التعامل مع الزبائن - جزء ٢	م. حنا بللوز	١٩٩٩
٣٥-تعلم كل شيء عن جافا	م.مهيب النقري	١٩٩٩
٣٦- مبادئ العمل السكرتاري باستخدام برنامج OUTLOOK	بيداء الزير	١٩٩٩
٣٧- أساسيات الإدارة المالية الحديثة	د. دريد درغام	١٩٩٩
٣٨- دليل التشخيص وتحديد الأهداف ووضع الخطط في المؤسسات	د. محمد مرعي مرعي	١٩٩٩
٣٩- التسويق وإدارة الأعمال التجارية	م. إياد زوكار	١٩٩٩
٤٠- أجهزة التحكم القابلة للبرمجة PLC	م. عبده هلاله	١٩٩٩
٤١- أمثلة وحالات عملية MS. EXCEL	م. إياد زوكار- م. نهال زركلي	٢٠٠٠
٤٢- المرجع الشامل لبرنامج 3D Studio Max - الجزء الثاني	م. جورج بركات	٢٠٠٠
٤٣- الأساليب الحديثة في التسويق	د. حسين علي	٢٠٠٠
٤٤- مرجع في صيانة الحواسب الشخصية	م. عبد الله أحمد	٢٠٠٠
٤٥- البرمجة في Access 2000	د. باسل الخطيب	٢٠٠٠
٤٦- دليل المحترفين إلى Corel Draw 9	م. سامر سعيد - م. حنان مسلّم - م. مصعب النقري	٢٠٠٠
٤٧- المرجع الشامل في برنامج معالجة النصوص MS Word 2000	د. هيثم البيطار - بوليت صارجي	٢٠٠٠

٤٨- مرجع أساسيات الحوسبة الجزء الأول: أساسيات الحاسوب	إشراف م.قاسم شعبان- شادي سيدا	٢٠٠٠
٤٩- دليل المديرين في إدارة الأفراد وفرق العمل	د. محمد مرعي مرعي	٢٠٠٠
٥٠- بناء التطبيقات باستخدام Oracle Developer	م. مهيب النقري	٢٠٠٠
٥١- فن وعلم إدارة الوقت	أ. رعد الصرن	٢٠٠٠
٥٢- الأخلاق الحديثة للإدارة الإدارة بالقيم	د. عدنان سليمان	٢٠٠٠
٥٣- من الفكرة إلى المنتج - إدارة الإبداع	د. حسين علي	٢٠٠٠
٥٤- دليل المطورين إلى دلفي Delphi	م. حسن شاليش حسن - م. سامر سعيد- م. ميشيل الياس	٢٠٠٠
٥٥-المعالجات التحكمية	م. عبده هلالة	٢٠٠٠
٥٦- الدليل العملي لتطبيق نظام الـ HACCP	م. ماهر العجي - م. ميلاد عريش	٢٠٠٠
٥٧- EXCEL 2000 - الجزء الأول	م. إياد زوكار- م. محمد الضمّاد	٢٠٠٠
٥٨- أساسيات الانترنت	د. ماهر سليمان- م. حسام عابد - م. إياد خدّام	٢٠٠٠
٥٩-الانترانت - بنيتها الأساسية وانعكاساتها على الشركات	د. عمار خير بك - م. حسام الملحم	٢٠٠٠

عناوين ستصدر قريباً

اسم الكتاب	المؤلف	تاريخ النشر المتوقع
١- التسويق عبر الانترنت	د. طلال عبود	٢٠٠٠
٢- نظام الشبكات WINDOWS NT	م.احسان مردود	٢٠٠٠
٣- تقنية المعلومات في إدارة الشركات	م. قاسم شعبان	٢٠٠٠
٤- تصميم المواقع WEB DESIGN	م.عبد الله أحمد	٢٠٠٠
٥- EXCEL 2000 - الجزء الثاني	م. إياد زوكار- م. محمد الضمّاد	٢٠٠٠
٦- المعلوماتية الطبية	د.نبيل دك الباب	٢٠٠٠
٧-أساسيات الإدارة المالية الحديثة - ج٢	د. دريد درشام	٢٠٠٠
٨- مرجع أساسيات الحوسبة ج٢	إشراف م. قاسم شعبان	
	شادي سيدا	٢٠٠٠
٩- المرجع الأساسي في		
MICROMEDIA DIRECTOR 7	أ. وائل جلال	٢٠٠٠
١٠- إدارة الابتكار والابداع	أ. رعد الصرن	٢٠٠٠
١١- تقنية المعلومات في		
إدارة الشركات	م. قاسم شعبان	٢٠٠٠
١٢ كتاب Autocad 2000	م. احسان مردود - م. وهبي معاد	٢٠٠٠
١٣- كتاب الحساسات وطرق		
الربط إلى أنظمة التحكم المبرمج	م. عبده هلاله - م. عامر عبود	٢٠٠٠
١٤- سلسلة الرضا لتبسيط علوم		
الحاسوب	م. مهيب النقري - د. معتصم شفا عمري	٢٠٠٠

سلسلة الرضا للمعلومات
دار الرضا للنشر

هذا الكتاب

أدى التطور في تكنولوجيا الاتصالات إلى انتشار المعلومات انتشاراً كبيراً، وبالتالي إلى الحاجة لأنظمة تقوم بالبحث الآلي عن المعلومات. وقد زاد في ذلك ظهور الانترنت ثم الشبكة العنكبوتية العالمية (World Wide Web "WWW") عليها، والتي تشكل ذروة الانفجار المعلوماتي في عصرنا هذا. فالعالم أجمع بتنوع معلوماته قد أصبح متاحاً لمختلف المستخدمين في جميع أنحاء العالم من خلال شاشة الحاسوب الصغيرة.

وأمام هذا الكم الهائل من المعلومات، يجب التفكير بحل جذري يسمح بالوصول إلى المعلومات التي يطلبها المستخدمون بسرعة وفاعلية.

مبتدئ متوسط متقدم

Cover by : Hot Design